



## Evaluation of room acoustic qualities and defects by use of auralization

**Rindel, Jens Holger**

*Published in:*  
Journal of Acoustical Society of America

*Publication date:*  
2004

*Document Version*  
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

*Citation (APA):*  
Rindel, J. H. (2004). Evaluation of room acoustic qualities and defects by use of auralization. In *Journal of Acoustical Society of America* (Vol. 116/4). Acoustical Society of America.

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# PROGRAM OF

## The 148th Meeting of the Acoustical Society of America

San Diego, California • Town and Country Hotel • 15–19 November 2004

1a MON. AM

NOTE: All Journal articles and Letters to the Editor are peer reviewed before publication. Program abstracts, however, are not reviewed before publication, since we are prohibited by time and schedule.

MONDAY MORNING, 15 NOVEMBER 2004

PACIFIC SALON 3, 8:25 A.M. TO 12:00 NOON

### Session 1aAA

#### Architectural Acoustics and Signal Processing in Acoustics: Topical Meeting on Spatial and Binaural Evaluation of Performing Arts Spaces I: Measurement Techniques and Binaural and Interaural Modeling

Ning Xiang, Cochair

*Rensselaer Polytechnic Institute, Architecture, 110 8th Street, Troy, New York 12180*

Rendell R. Torres, Cochair

*Rensselaer Polytechnic Institute, School of Architecture, 110 8th Street, Troy, New York 12180-3590*

Chair's Introduction—8:25

#### *Invited Papers*

8:30

**1aAA1. From manikin to microphone arrays development and application of binaural measurement devices.** Volker Mellert (Oldenburg Univ., Inst. f. Phys., 26111 Oldenburg, Germany, volker.mellert@uni-oldenburg.de)

Ever since stereophonic systems were designed it was obvious to make the recording by two microphone channels which mimic the aural investigation of the acoustic environment by our two ears. Head-related stereophony aims at a subjectively optimal reproduction of a sound field, in particular with headphones, whereas the multi-channel synthesis of a sound field provides a listening condition which is more independent of the individual listeners acoustic properties (e.g., head size, near-field refraction pattern). Binaural measurement devices are comparatively less complex and costly and therefore in use for about 30 years for sound field investigations, as in concert hall acoustics or in the assessment of environmental and technical sounds. A review is given on the development of head-related stereophony for investigating (mainly) room acoustics. Concepts of future devices for the assessment of technical sound are presented. The classical head-shaped recording system (dummy head) is substituted by beam-forming microphone arrays.

8:50

**1aAA2. Acoustic measurements through analysis of binaural recordings of speech and music.** David Griesinger (Harman Specialty Group, 3 Oak Park Dr., Bedford, MA 01730)

This paper will present and demonstrate some recent work on the measurement of acoustic properties from binaural recordings of live performances. It is found that models of the process of stream formation can be used to measure intelligibility, and, when combined with band-limited running cross-correlation, can be used to measure spaciousness and envelopment. Analysis of the running cross correlation during sound onsets can be used to measure the accuracy of azimuth perception. It is additionally found that the ease of detecting fundamental pitch from the upper partials of speech and music can be used as a measure of sound quality, particularly for solo instruments and singers.

9:10

**1aAA3. Some comparisons of binaural measurements made with different dummy heads and stereo microphone techniques.**

Peter A. Mapp (Peter Mapp Assoc., 5 Worthington Way, Copford, Colchester, Essex, UK)

Binaural measurements have been made in a number of acoustic environments, and the results from different binaural heads and stereo microphones are compared. The object of the study was not only to establish what practical differences occurred between the various head formats, but also to see if a stereo microphone or pseudohead could be used for making auditorium binaural measurements. Five measurement platforms were employed. These included two binaural dummy heads, binaural in-ear probe microphones, an SAAS pseudohead stereo microphone and a M-S (midside) stereo microphone. In the latter case, three different midside ratios were employed and compared. The measurements were made in a reverberant recital hall (2.5-s RT) and small acoustically treated listening room (RT 0.2 s). Whereas relatively minor differences were found to occur between the heads, significant differences were found to occur with the stereo microphones. It is concluded that while useful information can be obtained from a stereo microphone, it is far from being the same as binaural.

9:30

**1aAA4. Visualization of sound fields by three-dimensional representations.** Yasuhiro Tokita (RISE, Waseda Univ., Okubo 3-4-1, Shinjuku, 169-8555, Tokyo, Japan, tokita@acoust.rise.waseda.ac.jp) and Yoshio Yamasaki (Waseda Univ.)

In this study, three-dimensional sound fields can be visualized by measured data in an addition to transient numerical solutions. Concretely, the results of measurements by the four-point microphone method are used in the visualization of three-dimensional sound fields. By the results, distribution of virtual sources and hedgehog pattern are visualized as the receiving point is located on the center point in three-dimensional sound fields. The analyses of the transient sound fields are performed by the discretization of the wave equation. As the technique of numerical calculation, the finite difference method (FDM) is used. Transient sound fields are visualized by numerical solutions of sound pressure, instantaneous sound intensity, sound energy and particle displacement in three-dimensional space in addition to two-dimensional space. It has been verified that the numerical solutions have high accuracy by comparison with the corresponding analytical solutions. VRML97 is used as the viewer of three-dimensional sound fields. In this study, several trials of three-dimensional animation are going to be shown by VRML97 in order to visualize the transient sound fields.

9:50

**1aAA5. Directional measurements in sound fields using a spherical microphone array.** Bradford N. Gover (Inst. for Res. in Construction, Natl. Res. Council, Ottawa, ON K1A 0R6, Canada, brad.gover@nrc-cnrc.gc.ca)

A 32-element spherical array of microphones has been used to perform directional analysis of sound at a point. Optimized filter-and-sum beamforming is used to generate a highly directional beam over a broad frequency range. The beam can be steered in any three-dimensional direction, enabling angular discrimination of sound at a listening point. The directional responses can also be analyzed temporally, giving the directionality of the sound field at any time instant, or over any desired time range. By establishing a sound field with a test signal such as MLS or a sine sweep, the omnidirectional impulse response at each array microphone can be determined. Subsequent beamforming results in a set of directional impulse responses, which enables isolation of individual arrivals, or analysis in terms of early/late sound, for example. Various measurements made with the spherical array system will be presented and discussed.

10:10–10:20 Break

10:20

**1aAA6. Front-to-rear movement of reverberation in a concert hall.** Jonah Sacks and Mendel Kleiner (Program in Architectural Acoust., Rensselaer Polytechnic Inst., Troy, NY 12180)

In some concert halls one can perceive a front-to-rear movement of late reverberation as noted by T. J. Schultz [J. Sound Vib. **43**, 555–561 (1975)]. Listening tests show that the effect is not likely to be due to timbral change but rather to result from the spatial properties of the reverberant field. The effect can be noted as a change in late center time as a function of lateral angle of incidence of reverberant sound. The effect is likely to be a result of the scattering characteristics of hall surfaces. Measurements using a simplified scale model support this hypothesis. Measurements in Boston's Symphony Hall yield similar results. [Work supported by RPI.]

10:40

**1aAA7. Classification of binaural-activity patterns.** Wolfgang Hess and Jens Blauert (Inst. of Commun. Acoust., Ruhr-Universitaet, 44780 Bochum, Germany, wghess@gmx.de)

The three-dimensional, time-variant output of a computational model of the human auditory system was investigated with the aim of extracting perceptually patterns as relevant to spatial impression. To this end, the same stimuli with which the model was excited were also presented to human listeners which evaluated (i) the auditory source width (ASW), i.e., the horizontal extent of the auditory event, and (ii) the listener envelopment (LEV), i.e., the amount of envelopment by the auditory event. The listeners had to be trained beforehand as the discrimination of source- and room-related parameters of auditory events is not a common task to them. To allow for natural localization and to enable interactive listening, the stimuli were presented through head-tracked headphones. Trained listeners appeared to produce consistent results when depicting auditory objects on a compass-rose-shaped plane, where they had to mark the position and broadening of the auditory event(s) by the position(s) and length(s) of straight lines, and the envelopment by the position and size of an ellipse.

**1aAA8. Binaural signal analysis of diffuse sound fields.** Pedro A. Novo (Inst. of Commun. Acoust., IC1 132, Ruhr-Univ. of Bochum, 44780 Bochum, Germany, pedro.novo@ruhr-uni-bochum.de)

The simulation and the binaural recording of sounds produced by large crowds and rain impact sounds has produced an unexpected result. The listeners have reported that the auditory events were mainly concentrated to the left and to right, although the sound sources were equally distributed around the listeners. A similar result was reported with binaural recordings of applause. The results of a binaural signal analysis suggest that the key aspect regarding the lateral position dominance of the auditory events is connected to the sections where the cross-correlation coefficient assumes negative values. A comparison between normalized and non-normalized cross-correlation function predictions indicates that the latter is a better predictor for the cases studied. An adaptation period of several seconds was reported by several listeners. It is suggested that this adaptation period is related to the variations of the cross-correlation, which only average out after, approximately 2 s. The binaural impulse responses of a concert hall at two different listener positions have been analyzed in the light of the previous findings. In particular the diffuse sound field buildup will be analyzed in detail. Implications for the auditory source width and listener envelopment will be discussed.

**1aAA9. Auditory and visual spatial impression: Recent studies of three auditoria.** Andy Nguyen and Densil Cabrera (School of Architecture, Design Sci. and Planning, Wilkinson Bldg. G04, Univ. of Sydney, Sydney, NSW 2006, Australia)

Auditory spatial impression is widely studied for its contribution to auditorium acoustical quality. By contrast, visual spatial impression in auditoria has received relatively little attention in formal studies. This paper reports results from a series of experiments investigating the auditory and visual spatial impression of concert auditoria. For auditory stimuli, a fragment of an anechoic recording of orchestral music was convolved with calibrated binaural impulse responses, which had been made with the dummy head microphone at a wide range of positions in three auditoria and the sound source on the stage. For visual stimuli, greyscale photographs were used, taken at the same positions in the three auditoria, with a visual target on the stage. Subjective experiments were conducted with auditory stimuli alone, visual stimuli alone, and visual and auditory stimuli combined. In these experiments, subjects rated apparent source width, listener envelopment, intimacy and source distance (auditory stimuli), and spaciousness, envelopment, stage dominance, intimacy and target distance (visual stimuli). Results show target distance to be of primary importance in auditory and visual spatial impression—thereby providing a basis for covariance between some attributes of auditory and visual spatial impression. Nevertheless, some attributes of spatial impression diverge between the senses.

**1aAA10. Evaluation of a model of auditory source width based on the interaural cross-correlation coefficient.** Russell Mason, Tim Brookes, and Francis Rumsey (Inst. of Sound Recording, Univ. of Surrey, Guildford, Surrey, GU2 7XH, UK)

A model of human perception of auditory source width has been developed, based on the interaural cross-correlation coefficient (IACC). The following factors differentiate it from more commonly used IACC-based models: the use of a running analysis to quantify variations in width over time; half-wave rectification and low pass filtering of the input signal to mimic the breakdown of phase locking in the ear; compensation for the frequency and loudness dependency of perceived width; combination of a model of perceived location with a model of perceived width; and conversion of the results to an intuitive scale. Objective and subjective methods have been used to evaluate the accuracy and limitations of the resulting model.

## Session 1aBB

**Biomedical Ultrasound/Bioresponse to Vibration: Topical Meeting on Ultrasound Characterization of Cancellous and Cortical Bone I: Propagation and Scattering Properties**

Kendall R. Waters, Cochair

*National Institute of Standards and Technology, 325 Broadway, Boulder, Colorado 80305-3328*

Brent K. Hoffmeister, Cochair

*Department of Psychology, Rhodes College, 2000 North Parkway, Memphis, Tennessee 38112***Chair's Introduction—7:55*****Invited Papers*****8:00****1aBB1. Ultrasonic wave propagation and scattering in cancellous bone.** Keith A. Wear (FDA CDRH HFZ-142, 12720 Twinbrook Pky., Rockville, MD 20852)

Theoretical models and experimental data describing the interaction between ultrasound and cancellous bone will be discussed. Ultrasonic attenuation in cancellous bone is much greater than that for soft tissues and varies approximately linearly with frequency between 400 kHz and 1.7 MHz. Speed of sound in cancellous bone is slightly higher than that for soft tissues and decreases gradually with frequency at diagnostic frequencies (between 300 and 700 kHz). The dependence of phase velocity on porosity may be predicted from theory of acoustic propagation in fluid-filled porous solids. The negative dispersion can be explained using a stratified two-component model. At diagnostic frequencies, scattering varies approximately as frequency to the  $n$ th power where  $3 < n < 3.5$ . This may be explained either by a continuous or discrete scattering model. The latter represents trabeculae as small finite-length cylindrical scatterers. The frequency dependence of scattering, along with the anisotropy of scattering and attenuation, implies that absorption is a greater source of attenuation than scattering. [Funding from the FDA Office of Women's Health is gratefully acknowledged.]

**8:25****1aBB2. Ultrasonic wave propagation in human cancellous bone: Application of Biot theory.** Zine El Abidine Fellah, Jean Yves Chapelon (INSERM, unite 556, 151 cours Albert Thomas, 69424 Lyon), Walter Lauriks (Laboratorium voor Akoestiek en Thermische Fysica, Katholieke, France), and Claude Depollier (Laboratoire d'Acoustique de l'Universite du Maine, UMR-CNRS, France)

Ultrasonic wave propagation in human cancellous bone is considered. Reflection and transmission coefficients are derived for a slab of cancellous bone having an elastic frame using Biot's theory modified by the Johnson *et al.* model for viscous exchange between fluid and structure. Numerical simulations of transmitted waves in the time domain are worked out by varying the modified Biot parameters. The sensitivity of each physical parameter used in the theory has been studied in transmission. Some parameters play an important role in slow wave waveform, such as the viscous characteristic length and pore fluid bulk modulus. However, other parameters play an important role in the fast wave waveform, such as solid density and shear modulus. We also note from these simulations that some parameters, such as porosity, tortuosity, thickness, solid bulk modulus and skeletal compressibility frame, play an important role simultaneously in both fast and slow waveforms compared to other parameters which act on the waveform of just one of the two waves. Experimental results for slow and fast waves transmitted through human cancellous bone samples are given and compared with theoretical predictions.

**8:50****1aBB3. Simulation of ultrasound propagation in bone.** Jonathan J. Kaufman, Gangming Luo (CyberLogic, Inc., 611 Broadway, Ste. 707, New York, NY 10012, jkkaufman@cyberlogic.org), and Robert S. Siffert (Mount Sinai School of Medicine, New York, NY 10012)

Ultrasound has been proposed as a means to noninvasively assess bone and, particularly, bone strength and fracture risk, as for example in osteoporosis. Because strength is a function of both mineral density and architecture, ultrasound has the potential to provide more accurate measurement of bone integrity than, for example, with x-ray absorptiometric methods. Although some of this potential has already been realized—a number of clinical devices are presently available—there is still much that is unknown regarding the interaction of ultrasound with bone. Because of the inherent complexity of the propagation medium, few analytic solutions exist with practical application. For this reason, ultrasound simulation techniques have been developed and applied to a number of different problems of interest in ultrasonic bone assessment. Both 2D and 3D simulation results will be presented, including the effects of architecture and density on the received waveform, propagation effects of both cortical and trabecular bone, and the

## Contributed Papers

9:15

**1aBB4. Apparent integrated backscatter from cancellous bone in the frequency range 2.5–7.5 MHz.** Brent K. Hoffmeister, Charles I. Jones III (Dept. of Phys., Rhodes College, 2000 North Pkwy., Memphis, TN 38112, hoffmeister@rhodes.edu), and Sue C. Kaste (St. Jude Children's Res. Hospital, Memphis, TN 38105)

Ultrasonic backscatter may offer useful new techniques for ultrasonic bone assessment. For this study, backscatter measurements were performed on 18 specimens of bovine cancellous bone using a 5-MHz broadband ultrasonic system operating between 2.5–7.5 MHz. Specimens were obtained from the proximal tibia, and prepared in the shape of cubes (15-mm side length) with faces oriented along principal anatomic directions (anterior, posterior, medial, lateral, superior, and inferior). A mechanical scanning system was used to acquire ultrasonic backscatter signals from 144 sites perpendicular to each face of each cube. The signals were analyzed to extract values of apparent integrated backscatter (AIB), a parameter that represents the frequency-averaged backscattered power uncompensated for the effects of attenuation and diffraction. AIB demonstrated highly significant linear correlations with bone mineral density (BMD) for both the transverse (anterior, posterior, medial, and lateral) and longitudinal (superior and inferior) directions. In all cases AIB decreased with increasing BMD. No significant anisotropy was detected. We conclude that AIB measurements performed on bovine cancellous bone in this range of frequencies correlate well with BMD.

9:30

**1aBB5. Some aspects of the atypical acoustic behavior of cancellous bone.** Patrick H. Nicholson (Dept. of Health Sci., Univ. of Jyväskylä, P.O. Box 35, 40014, Finland, patrick\_nicholson@btopenworld.com)

Cancellous bone, consisting of a porous framework of solid trabeculae saturated in marrow fluid, is a challenging material in terms of ultrasonic characterization. This study considers some unusual aspects of the ultrasonic behavior of cancellous bone. These anomalies highlight the boundaries of our knowledge and, arguably, point us towards promising directions for future progress. Profoundly different wave propagation regimes can exist along different anatomical axes. Along the main axis of trabecular orientation, two longitudinal waves are observed, but only one wave is observed for propagation transversely across the trabeculae. In the latter case, a positive attenuation slope and negative velocity dispersion are seen, in apparent violation of the Kramers–Kronig relations. The properties of the saturating fluid have major effects on acoustic properties, and temperature studies also point towards the importance of the fluid. In some clinical studies, velocity and attenuation in the heel show opposite changes, which is difficult to explain without considering the possibility of changes in the marrow as well as the bone. Shear wave propagation in fluid-filled cancellous bone has received little attention. Measurements in whale bone suggest that shear waves propagate in dry bone, but are rapidly attenuated in the presence of saturating fluid.

9:45

**1aBB6. Dependence of dispersive properties of cancellous bone on bone mineral and collagen content.** K. R. Waters (NIST, Mater. Reliability Div., 325 Broadway, Boulder, CO 80305), B. K. Hoffmeister, and S. A. Whitten (Rhodes College, Memphis, TN 38112)

Cancellous bone is known to exhibit dispersion over clinically-relevant ultrasonic frequencies. However, the dependence of the dispersive properties of cancellous bone on bone mineral and collagen content are currently not well understood. Cancellous bone specimens ( $n=16$ ) obtained from the proximal end of bovine tibia were cut into cubes (15 mm sides) and prepared for measurement along the anterior-posterior (AP), medial-lateral

(ML), and superior-inferior (SI) directions. Bone mineral densities (BMD) ranged from 100 to 250 kg/m<sup>3</sup>. Subsets of the bone specimens were chemically treated to remove bone mineral ( $n_M=10$ ) and collagen content ( $n_C=6$ ). Ultrasonic measurements were performed before and after treatment in a water bath (23 °C) using a pair of 2.25-MHz planar transducers aligned coaxially. Phase velocity was determined from Fourier analysis of the ultrasonic signals, and dispersion was calculated as the change in phase velocity across the usable bandwidth. Demineralization produced highly significant changes in dispersion along the AP ( $p=0.004$ ) and ML ( $p=0.002$ ) directions, but not the SI direction ( $p=0.6$ ). Decollagenization produced no significant change for any direction ( $p>0.3$ ). These results indicate that bone mineral plays a much larger role than collagen in the dispersive behavior of cancellous bone.

10:00–10:15 Break

10:15

**1aBB7. Ultrasound attenuation as a quantitative measure of fracture healing.** Sabina Gheduzzi (Orthopaedic Surgery, Univ. of Bristol, AOC Lower Level, Southmead Hospital, Bristol BS10 5NB, UK, s.gheduzzi@bristol.ac.uk), Victor F. Humphrey (Univ. of Southampton, UK), Simon P. Dodd, James L. Cunningham, and Anthony W. Miles (Univ. of Bath, UK)

The monitoring of fracture healing still relies upon the judgment of callus formation and on the manual assessment of the stiffness of the fracture. A diagnostic tool capable of quantitatively measuring healing progression of a fracture would allow the fine-tuning of the treatment regime. Ultrasound attenuation measurements were adopted as a possible method of assessing the healing process in human long bones. The method involves exciting ultrasonic waves at 200 kHz in the bone and measuring the reradiation along the bone and across the fracture zone. Seven cadaveric femora were tested *in vitro* in intact form and after creating a transverse fracture by sawing through the cortex. The effects of five different fracture types were investigated. A partial fracture, corresponding to a 50% cut through the cortex, a closed fracture, and fractures of widths varying between 1, 2, and 4 mm were investigated. The introduction of a fracture was found to produce a dramatic effect on the amplitude of the signal. Ultrasound attenuation was found to be sensitive to the presence of a fracture, even when the fracture was well reduced. It would therefore appear feasible to adopt attenuation across a fracture as a quantitative measurement of fracture healing.

10:30

**1aBB8. Experimental evaluation of bone quality using speed of sound measurement in cadaver mandibles.** Iyad Al Haffar, Frederic Padilla, Pascal Laugier (Laboratoire d'Imagerie Paramétrique, CNRS UMR 7623, Université Paris 6, 15 rue de l'Ecole de Médecine, 75006 Paris, France), Raphael Nefussi (Laboratoire de Biologie, 75006 Paris, France), Sami Kolta (Université Ren Descartes, Paris, France), and Jean-Michel Foucart (Service d'Imagerie Médicale Hotel-Dieu de Paris)

This study is the first attempt to use speed of sound (SOS) as a new ultrasonic diagnostic tool for bone quality assessment before oral implant treatment. The objective is to demonstrate the *in vitro* feasibility of local SOS measurement at the mandible, and to investigate the relationships between mandibular SOS and local bone mineral density (BMD) and the ratio between the trabecular and cortical thickness (Tb.Th/Cort.Th). Fourteen excised human mandible were measured in transmission with a pair of flat 1.6-MHz central frequency transducers. Three regions of interest (ROIs) were selected in the specimens: incisor, premolar and molar regions. Ten measurements with repositioning were performed on each ROI

to determine the short-term precision. Dual x-ray absorptiometry scans were performed on the samples for local BMD measurements. Computed tomography (CT) was used to determine mandibular cross-sectional morphological measurements. SOS measurements at different sites were significantly different, reflecting the heterogeneity between the different sites. A strong linear relationship was found between SOS and BMD ( $r^2 = 0.68$ ,  $p < 0.0001$ ) while a nonlinear relationship was found between SOS and Tb.Th/Cort.Th ( $r^2 = 0.48$ ,  $p < 0.0001$ ). This study demonstrates *in vitro* the feasibility of SOS measurement at the mandible. *In vitro* mandibular SOS reflects local BMD and Tb.Th/Cort.Th before implant.

10:45

**1aBB9. Investigation of ultrasonic properties of human hip bone: Relationship to bone mineral density (BMD) and micro-architecture.**

Frederic Jenson, Frederic Padilla, and Pascal Laugier (Laboratoire d'Imagerie Paramétrique, CNRS UMR 7623 Université Paris 6, 15 rue de l'Ecole de Médecine, 75006 Paris, France)

The goal of this study is to investigate the relationships between QUS measurements (transmission and reflection) and BMD and micro-architecture measured on human proximal femur. Thirty-eight 1-cm thick slices of trabecular bones were removed from fresh human proximal femurs. Two-dimensional scans were performed using 1-MHz focused transducers to measure QUS parameters: normalized BUA (nBUA), SOS and broadband ultrasonic backscatter (BUB). BMD was determined using QCT. Six hundred sixty non-overlapping ROIs were selected for quantitative analysis on both QUS and QCT images. Finally, 37 cylindrical cores were extracted from 31 specimens (8 mm diameter). Their micro-architecture was derived using micro-tomography. All QUS parameters were significantly correlated to BMD (nBUA:  $r^2 = 0.7$ ; SOS:  $r^2 = 0.75$ ; BUB:  $r^2 = 0.44$ ). Multiple regression models were tested with BMD and micro-architecture. SOS only was best explained by a combination of bone quantity (BV/TV) and micro-architecture parameters (SOS with BMD and structure model index:  $r^2 = 0.93$ ). No additional variability of nBUA and BUB could be explained by micro-architectural parameters in addition to BMD or BV/TV. These results suggest that QUS measurement at the hip may be relevant for fracture risk prediction.

11:00

**1aBB10. Numerical simulation of wave propagation in cancellous bone.** Emmanuel Bossy (Dept. of Aerosp. and Mech. Eng., Boston Univ., Boston, MA 02215), Frederic Padilla, Frederic Jenson, Guillaume Haiat, and Pascal Laugier (Laboratoire d'Imagerie Paramétrique, 75006 Paris, France)

Physical mechanisms implied in the propagation of ultrasound waves in trabecular bones are not yet clearly understood. The relative role of scattering and absorption in the attenuation process is unknown. To dissociate these two effects, a simulation of wave propagation in 3D volumes of trabecular bone is performed, and compared to experimental results. The simulation algorithm accounts for scattering and volumetric absorption into both the saturating fluid and bone. The simulation software uses a finite difference approach based on the Virieux numerical scheme. An incident plane wave is propagated on a volume of bone of approximately  $5 \times 5 \times 8 \text{ mm}^3$ . These volumes were reconstructed from high resolution micro-CT experiments. Several types of simulations are computed: bone described as a fluid matrix saturated by fluid (i.e., neglecting the shear wave propagation into the matrix) and bone described as a solid matrix saturated by fluid, with or without absorption. From the transmitted signals, attenuation and speed of sound through the specimens are calculated and compared to experimental values obtained on the same specimens. Finally, the contribution of scattering to the total attenuation is estimated, and the importance of mode conversions as well as absorption.

11:15

**1aBB11. Ultrasonic wave propagation in cortical bone mimics.** Simon P. Dodd, James L. Cunningham, Anthony W. Miles (Ctr. for Orthopaedic Biomechanics, Univ. of Bath, Bath, BA2 7AY, UK), Victor F. Humphrey (Univ. of Southampton, Southampton, SO17 1BJ, UK), and Sabina Gheduzzi (Univ. of Bristol, Bristol, BS10 5NB, UK)

Understanding the velocity and attenuation of ultrasonic waves in cortical bone is important for studies of osteoporosis and fractures. In particular, propagation in free- and water-loaded acrylic plates, with a thickness range of around 1–6 mm, has been widely used to mimic cortical bone behavior. A theoretical investigation of Lamb mode propagation at 200 kHz in free- and water-loaded acrylic plates revealed a marked difference in the form of their velocity and attenuation dispersion curves as a function of frequency thickness product. In experimental studies, this difference between free and loaded plates is not seen. Over short measurement distances, the results for both free and loaded plates are consistent with previous modeling and experimental studies: for thicker plates (above 3–4 mm), the velocity calculated using the first arrival signal is a lateral wave comparable with the longitudinal velocity. As the plate thickness decreases, the velocity approaches the  $S_0$  Lamb mode value. WAVE2000 modeling of the experimental setup agrees with experimental data. The data are also used to test a hypothesis that for thin plates the velocity approaches the corresponding  $S_0$  Lamb mode velocity at large measurement distances or when different arrival time criteria are used. [Work supported by Action Medical Research.]

11:30

**1aBB12. Structural and elastic determinants of axial transmission ultrasonic velocity in the human radius.** Kay Raum (Q-BAM Group, Dept. of Orthopedics, Martin Luther Univ., 06097 Halle, Germany; kay.raum@medizin.uni-halle.de), Ingrid Leguérney, Florent Chandelier, Maryline Talmant, Amena Saied, Pascal Laugier (Laboratoire d'Imagerie Paramétrique, 75006 Paris, France), and Françoise Peyrin (CREATIS and European Synchrotron Radiation Facility, 38043 Grenoble, France)

Accurate clinical interpretation of the sound velocity derived from axial transmission devices requires a detailed understanding of the propagation phenomena involved and of the bone factors that have an impact on measurements. In the low-megahertz range, ultrasonic propagation in cortical bone depends on anisotropic elastic tissue properties, porosity, and the spatial dimensions, e.g., cortical thickness. A subset of ten human radius samples from a previous biaxial transmission investigation was inspected using 50-MHz scanning acoustic microscopy (SAM) and synchrotron radiation computed tomography (SR-CT). Low-frequency axial transmission sound speed at 1 and 2 MHz was related to structural properties (cortical thickness C.Th, porosity POR, Haversian cavity density CDH) and tissue parameters (acoustic impedance Z, mineral density MD) on site-matched cross sections. Significant linear multivariate regression models (1 MHz:  $R = 0.84$ ,  $p < 1E-4$ , 2 MHz:  $R = 0.65$ ,  $p < 1E-4$ ) were found for the combination of C.Th with POR and Z (measured in the external cortical quarter). A modified model accounting for the nonlinear dispersion relation with C.Th was also highly significant ( $R = 0.75$ ,  $p < 1E-4$ ,  $\text{rmse} = 49.22 \text{ m/s}$ ) and explained (after adjustment for dispersion) 55.6% of the variance of the sound velocity by variations of porosity (15.6%) and impedance (40%).

11:45

**1aBB13. Simulated axial transmission propagation on cortical bone.** M. Talmant, E. Bossy, and P. Laugier (Laboratoire d'Imagerie Paramétrique, UMR CNRS 7623, 15 rue de l'école de Médecine, 75006 Paris, France)

The ultrasonic axial transmission technique, used to assess cortical shells of long bones, is investigated using numerical simulations. The 3-D finite difference code generates synthetic signals recorded at different dis-

tance emitter–receiver for a linear arrangement of transducers placed along the bone axis. Academic modeling of bone based on cylindrical tubular shape made of anisotropic and porous material has been reported [J. Acoust. Soc. Am. **115** (2004)]. The aim of this paper is to refine the model by taking into account more realistic structural and material bone properties. Finite difference modeling was applied to 50 human radius specimens which were examined both by x-ray tomography at different resolutions and by ultrasonic axial transmission technique (1 MHz). The x-ray macroscopic geometry (pixel around 100  $\mu\text{m}$ ) of the 50 samples

was imported in the code. Material properties are assigned to each bone according to its own microarchitecture examined at a 10  $\mu\text{m}$  scale. Different assumptions of the relationship between bone structural properties and material properties were tested. Simulations were validated by comparison with experimental results. Numerical simulations of transient propagation in bone is a powerful tool to enlighten interaction between ultrasound and bone and consequently to improve ultrasound based devices for clinical use.

MONDAY MORNING, 15 NOVEMBER 2004

CALIFORNIA ROOM, 9:00 A.M. TO 12:00 NOON

## Session 1aSC

### Speech Communication and Signal Processing in Acoustics: Current Developments in Speech Communication: Techniques and Technology (Poster Session)

Shrikanth Narayanan, Chair

*Signal and Image Processing Institute, University of Southern California, 3740 McClintock Avenue, Los Angeles, California 90089-2564*

#### Contributed Papers

All posters will be on display from 9:00 a.m. to 12:00 noon. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 9:00 a.m. to 10:30 a.m. and contributors of even-numbered papers will be at their posters from 10:30 a.m. to 12:00 noon.

**1aSC1. Supplementary features for improving phone recognition.** Mridul Balaraman, Sorin Dusan, and James L. Flanagan (Ctr. for Adv. Information Processing, 96 Frelinghuysen Rd., Piscataway, NJ 08855, mridulb@caip.rutgers.edu)

Traditional speech recognition systems use mel-frequency cepstral coefficients (MFCCs) as acoustic features. The present research aims to study the classification characteristics and the performance of some supplementary features (SFs) such as periodicity, zero crossing rate, log energy and ratio of low frequency energy to total energy, in a phone recognition system, built using the Hidden Markov Model Tool Kit. To demonstrate the performance of the SFs, training is done on a subset of the TIMIT data base (DR1 data set) on context independent phones using a single mixture. When only the SFs and their first derivatives (feature set of dimension 8) are used the recognition accuracy is found to be 42.96% as compared to 54.65% when 12 MFCCs and their corresponding derivatives are used. The performance of the system improves to 56.49%, when the SFs and their derivatives are used along with the MFCCs. A further improvement to 60.34% is observed when the last 4 MFCCs and their derivatives are replaced by SFs and their derivatives, respectively. These results indicate that the supplementary features contain classification characteristics which can be useful in automatic speech recognition.

**1aSC2. Non-monotonic spectral transitions between successive phonemes.** Sorin Dusan (Ctr. for Adv. Information Processing, Rutgers Univ., 96 Frelinghuysen Rd., Piscataway, NJ 08854, sdusan@caip.rutgers.edu)

In fluent speech one would expect that the transition between two successive acoustic targets would exhibit either a monotonic increasing or a monotonic decreasing trajectory in the time domain of each spectral parameter. Closer examination reveals that this is not always the case. This study investigates the existence of non-monotonic trajectories found in the acoustic spectral domain at the transition between some successive phonemes. This non-monotonic behavior consists of transitional regions that

exhibit either increasing and then decreasing or decreasing and then increasing trajectories for some acoustic spectral parameters. Using the superposition principle, the training part of the TIMIT acoustic-phonetic database is used to build 2471 diphone trajectory models, based on the 61 symbols used in the database for phonetic transcription. Various non-monotonic trajectories are found in some of these models for some spectral representations including the linear prediction coding (LPC) parameters and the mel-frequency cepstral coefficients (MFCC). In some cases these spectral peaks or valleys between phonemes have significant amplitudes reaching approximately the standard deviation at the center, positions of the adjacent phonemes. These non-monotonic trajectories cannot be explained by the Stevens quantal theory of speech in which abrupt changes appear at transitions between two acoustic targets.

**1aSC3. Learning segments from waveforms: A statistical model of early phonological acquisition.** Ying Lin (Phonet. Lab, Linguist. Dept., UCLA, Los Angeles, CA 90095-1543)

We consider the problem of early phonological acquisition within a statistical learning framework. Given a set of waveforms, our goal is to output a set of models for phonological categories as well as representations of the utterances in terms of these categories. The basic assumption of our model is that the learning of sound categories neither precedes nor follows segmentation of the waveform, but improves together with segmentation in an iterative manner. In statistical learning terms, segmentation is regarded as “missing data,” while category models are treated as parameters of interest. Our algorithm iterates over two basic steps: First, given a set of category models, we compute a mixture of segmentations for each utterance; Second, we improve the category models with unlabeled segments using unsupervised learning. Starting from an acoustic segmentation, each iteration produces an initial estimate of the parameters for the next iteration, until the algorithm converges to a set of sublexical models and a set of segmentations for each utterance. The results of run-



ning this algorithm on TIMIT and motherese data suggest that it approximately identifies segment-sized units and their associated categories in a completely unsupervised manner.

#### **1aSC4. Phoneme recognition with kernel learning algorithms.**

Hassan H. Namarvar and Theodore W. Berger (Dept. of Biomed. Eng., Univ. of Southern California, OHE-500, Los Angeles, CA 90089-1451, heidarin@usc.edu)

An isolated phoneme recognition system is proposed using time-frequency domain analysis and support vector machines (SVMs). The TIMIT corpus which contains a total of 6300 sentences, ten sentences spoken by each of 630 speakers from eight major dialect regions of the United States, was used in this experiment. Provided time-aligned phonetic transcription was used to extract phonemes from speech samples. A 55-output classifier system was designed corresponding to 55 classes of phonemes and trained with the kernel learning algorithms. The training dataset was extracted from clean training samples. A portion of the database, i.e., 65 338 samples of training dataset, was used to train the system. The performance of the system on the training dataset was 76.4%. The whole test dataset of the TIMIT corpus was used to test the generalization of the system. All samples, i.e., 55 655 samples of the test dataset, were used to test the system. The performance of the system on the test dataset was 45.3%. This approach is currently under development to extend the algorithm for continuous phoneme recognition. [Work supported in part by grants from DARPA, NASA, and ONR.]

**1aSC5. Spoken name pronunciation evaluation.** Joseph Tepperman and Shrikanth Narayanan (Speech Anal. and Interpretation Lab., USC, 3740 McClintock Ave., Ste. 131, Los Angeles, CA 90089, tepperma@usc.edu)

Recognition of spoken names is an important ASR task since many speech applications can be associated with it. However, the task is also among the most difficult ones due to the large number of names, their varying origins, and the multiple valid pronunciations of any given name, largely dependent upon the speaker's mother tongue and familiarity with the name. In order to explore the speaker- and language-dependent pronunciation variability issues present in name pronunciation, a spoken name database was collected from 101 speakers with varying native languages. Each speaker was asked to pronounce 80 polysyllabic names, uniformly chosen from ten language origins. In preliminary experiments, various prosodic features were used to train Gaussian mixture models (GMMs) to identify misplaced syllabic emphasis within the name, at roughly 85% accuracy. Articulatory features (voicing, place, and manner of articulation) derived from MFCCs were also incorporated for that purpose. The combined prosodic and articulatory features were used to automatically grade the quality of name pronunciation. These scores can be used to provide meaningful feedback to foreign language learners. A detailed description of the name database and some preliminary results on the accuracy of detecting misplaced stress patterns will be reported.

**1aSC6. Determining the threshold for usable speech within co-channel speech with the SPHINX automated speech recognition system.** William T. Hicks and Robert E. Yantorno (Speech Processing Lab., Elec. and Comput. Eng. Dept., College of Eng., Temple Univ., Philadelphia, PA 19122, bill.hicks@ieee.org)

Much research has been and is continuing to be done in the area of separating the original utterances of two speakers from co-channel speech. This is very important in the area of automated speech recognition (ASR), where the current state of technology is not nearly as accurate as human listeners when the speech is co-channel. It is desired to determine what types of speech (voiced, unvoiced, and silence) and at what target to interference ratio (TIR) two speakers can speak at the same time and not reduce speech intelligibility of the target speaker (referred to as usable

speech). Knowing which segments of co-channel speech are usable in ASR can be used to improve the reconstruction of single speaker speech. Tests were performed using the SPHINX ASR software and the TIDIGITS database. It was found that interfering voiced speech with a TIR of 6 dB or greater (on a per frame basis) did not significantly reduce the intelligibility of the target speaker in co-channel speech. It was further found that interfering unvoiced speech with a TIR of 18 dB or greater (on a per frame basis) did not significantly reduce the intelligibility of the target speaker in co-channel speech.

#### **1aSC7. Noise-robust speech recognition using multi-band spectral features.**

Yoshitaka Nishimura, Takahiro Shinozaki, Koji Iwano, and Sadaoki Furui (Tokyo Inst. of Technol., Dept. of Comput. Sci., 2-12-1 Ookayama, Meguro-ku, Tokyo, 152-8552, Japan)

In most of the state-of-the-art automatic speech recognition (ASR) systems, speech is converted into a time function of the MFCC (Mel Frequency Cepstrum Coefficient) vector. However, the problem with using the MFCC is that noise effects spread over all the coefficients even when the noise is limited within a narrow frequency band. If a spectrum feature is directly used, such a problem can be avoided and thus robustness against noise could be expected to increase. Although various researches on using spectral domain features have been conducted, improvement of recognition performances has been reported only in limited noise conditions. This paper proposes a novel multi-band ASR method using a new log-spectral domain feature. In order to increase the robustness, log-spectrum features are normalized by applying the three processes: subtracting the mean log-energy for each frame, emphasizing spectral peaks, and subtracting the log-spectral mean averaged over an utterance. Spectral component likelihood values in each frame are weighted by normalized spectral level (spectral peaks) or SNR of each component. Experimental results using babble noise-added speech show that recognition performance is improved by the proposed method in comparison with the MFCC-based method. The performance is further improved by spectral-peak weighting and SNR-based frequency-band weighting techniques.

#### **1aSC8. On audio-visual synchronization for viseme-based speech synthesis.**

Jianxia Xue, Abeer Alwan (Dept. of Elec. Eng., UCLA, 405 Hilgard Ave., Los Angeles, CA 90095), Edward T. Auer, Jr., and Lynne E. Bernstein (House Ear Inst., Los Angeles, CA 90057)

In viseme-based visual speech synthesis [e.g., T. Ezzat and T. Poggio, *Int. J. Comput. Vision* **38**, 45–57 (2000)], it is frequently assumed that each viseme can be represented by an image which corresponds to the temporal mid-point of the phoneme's acoustic segment. We investigated the validity of this assumption in a study of 3-D motion data extracted from marker positions on the face of one talker. Using principle components analysis (PCA), the first five principle motion tracks comprising the lips and the jaw were studied. The temporal positions of the local extrema in the principle motion tracks were analyzed. Statistics of the distances between these local extrema and the mid-points of the corresponding acoustic segments were compared. Results showed that the local extrema of the motion tracks were, for the most part, not well-aligned to the phoneme's acoustic midpoint with a few exceptions. For example, for /s/ the local extrema were well aligned (small variance), while for /m/ and /f/, alignments pattern could be found for several, but not all, tokens. The results suggest that phonetic context and speaking rate must be taken into account when characterizing facial configurations in visual speech synthesis. [Work supported in part by the NSF.]

**1aSC9. Emotion to emotion speech conversion in phoneme level.**

Murtaza Bulut, Serdar Yildirim, Carlos Busso, Chul Min Lee, Ebrahim Kazemzadeh, Sungbok Lee, and Shrikanth Narayanan (Univ. of Southern California, Los Angeles, CA 90089)

Having an ability to synthesize emotional speech can make human-machine interaction more natural in spoken dialogue management. This study investigates the effectiveness of prosodic and spectral modification in phoneme level on emotion-to-emotion speech conversion. The prosody modification is performed with the TD-PSOLA algorithm (Moulines and Charpentier, 1990). We also transform the spectral envelopes of source phonemes to match those of target phonemes using LPC-based spectral transformation approach (Kain, 2001). Prosodic speech parameters ( $F_0$ , duration, and energy) for target phonemes are estimated from the statistics obtained from the analysis of an emotional speech database of happy, angry, sad, and neutral utterances collected from actors. Listening experiments conducted with native American English speakers indicate that the modification of prosody only or spectrum only is not sufficient to elicit targeted emotions. The simultaneous modification of both prosody and spectrum results in higher acceptance rates of target emotions, suggesting that not only modeling speech prosody but also modeling spectral patterns that reflect underlying speech articulations are equally important to synthesize emotional speech with good quality. We are investigating supra-segmental level modifications for further improvement in speech quality and expressiveness.

**1aSC10. Study of acoustic correlates associate with emotional speech.**

Serdar Yildirim, Sungbok Lee, Chul Min Lee, Murtaza Bulut, Carlos Busso, Ebrahim Kazemzadeh, and Shrikanth Narayanan (Univ. of Southern California, Los Angeles, CA 90089)

This study investigates the acoustic characteristics of four different emotions expressed in speech. The aim is to obtain detailed acoustic knowledge on how a speech signal is modulated by changes from neutral to a certain emotional state. Such knowledge is necessary for automatic emotion recognition and classification and emotional speech synthesis. Speech data obtained from two semi-professional actresses are analyzed and compared. Each subject produces 211 sentences with four different emotions; neutral, sad, angry, happy. We analyze changes in temporal and acoustic parameters such as magnitude and variability of segmental duration, fundamental frequency and the first three formant frequencies as a function of emotion. Acoustic differences among the emotions are also explored with mutual information computation, multidimensional scaling and acoustic likelihood comparison with normal speech. Results indicate that speech associated with anger and happiness is characterized by longer duration, shorter interword silence, higher pitch and rms energy with wider ranges. Sadness is distinguished from other emotions by lower rms energy and longer interword silence. Interestingly, the difference in formant pattern between [happiness/anger] and [neutral/sadness] are better reflected in back vowels such as /a/(father/) than in front vowels. Detailed results on intra- and interspeaker variability will be reported.

**1aSC11. Analysis and synthesis of laughter.** Shiva Sundaram and Shrikanth Narayanan (Speech Anal. and Interpretation Lab., Univ. of Southern California, 3740 McClintock Ave., EEB400, Los Angeles, CA 90089, ssundara@usc.edu)

There is much enthusiasm in the text-to-speech community for synthesis of emotional and natural speech. One idea being proposed is to include emotion dependent paralinguistic cues during synthesis to convey emotions effectively. This requires modeling and synthesis techniques of various cues for different emotions. Motivated by this, a technique to synthesize human laughter is proposed. Laughter is a complex mechanism of expression and has high variability in terms of types and usage in human-human communication. People have their own characteristic way of laughing. Laughter can be seen as a controlled/uncontrolled physiological process of a person resulting from an initial excitation in context. A parametric model based on damped simple harmonic motion to effectively

capture these diversities and also maintain the individuals characteristics is developed here. Limited laughter/speech data from actual humans and synthesis ease are the constraints imposed on the accuracy of the model. Analysis techniques are also developed to determine the parameters of the model for a given individual or laughter type. Finally, the effectiveness of the model to capture the individual characteristics and naturalness compared to real human laughter has been analyzed. Through this the factors involved in individual human laughter and their importance can be better understood.

**1aSC12. Effects of emotion on different phoneme classes.**

Chul Min Lee, Serdar Yildirim, Murtaza Bulut, Carlos Busso, Abe Kazemzadeh, Sungbok Lee, and Shrikanth Narayanan (Univ. of Southern California, Los Angeles, CA 90089)

This study investigates the effects of emotion on different phoneme classes using short-term spectral features. In the research on emotion in speech, most studies have focused on prosodic features of speech. In this study, based on the hypothesis that different emotions have varying effects on the properties of the different speech sounds, we investigate the usefulness of phoneme-class level acoustic modeling for automatic emotion classification. Hidden Markov models (HMM) based on short-term spectral features for five broad phonetic classes are used for this purpose using data obtained from recordings of two actresses. Each speaker produces 211 sentences with four different emotions (neutral, sad, angry, happy). Using the speech material we trained and compared the performances of two sets of HMM classifiers: a generic set of "emotional speech" HMMs (one for each emotion) and a set of broad phonetic-class based HMMs (vowel, glide, nasal, stop, fricative) for each emotion type considered. Comparison of classification results indicates that different phoneme classes were affected differently by emotional change and that the vowel sounds are the most important indicator of emotions in speech. Detailed results and their implications on the underlying speech articulation will be discussed.

**1aSC13. Representation for dialect recognition using topographic independent component analysis.**

Qu wei (Dept. of Information Technol., Univ. of Sci. and Technol. of Beijing in China, 100084, Anhuili, Chaoyang District, Beijing, China)

In dialect speech recognition, the feature of tone in one dialect is subject to changes in pitch frequency as well as the length of tone. It is beneficial for the recognition if a representation can be derived to account for the frequency and length changes of tone in an effective and meaningful way. In this paper, we propose a method for learning such a representation from a set of unlabeled speech sentences containing the features of the dialect changed from various pitch frequencies and time length. Topographic independent component analysis (TICA) is applied for the unsupervised learning to produce an emergent result that is a topographic matrix made up of basis components. The dialect speech is topographic in the following sense: the basis components as the units of the speech are ordered in the feature matrix such that components of one dialect are grouped in one axis and changes in time windows are accounted for in the other axis. This provides a meaningful set of basis vectors that may be used to construct dialect subspaces for dialect speech recognition.

**1aSC14. Robust speaker recognition using glottal information-based cepstral mean subtraction.**

Pu Yang, Yingchun Yang, and Zhaohui Wu (Zheda Rd. 38#, Hangzhou, Zhejiang Province, Peoples' Republic of China)

Channel distortion and background noise often severely degrade the performance of automatic speaker recognition (ASR) system. In this paper, a new compensation method called glottal information-based cepstral mean subtraction (GIBCMS), which improves upon the conventional cepstral mean subtraction (CMS) method, is presented. Besides the cepstral

information, GIBCMS has utilized the speaker's glottal information, which also holds speaker-dependent characteristics, but is less vulnerable to environment than the cepstral one. In order to test its robustness under channel distortion, even with high level of background noise, we applied this method to the SRMC corpus which is added by noise at different SNRs. The experimental results show that GIBCMS gains better performance over other improved CMS methods on it.

**1aSC15. Systematic speaker variation and within-speaker center of gravity correlations in the TIMIT database.** Hector R. Javkin, Carol Christie, Gaston R. Cangiano, Elaine Drom (Dept. of Linguist. and Lang. Development, San Jose State Univ., San Jose, CA 95192-0093), and Katia McClain (Univ. of California, Santa Barbara, CA)

The systematicity of speaker variation in consonants was examined by measuring the noise component of consonants in the TIMIT database. Fricatives and stops were compared by measuring at the temporal middle of fricatives and at the release of stops. There is a high correlation between many of the center of gravity measures. For example, if a speaker has a particularly high center of gravity in the sound /t/, she or he will also have a high center of gravity in /d/, /s/, /p/, /k/, and other consonants. The full set of correlations will be described in the paper. The correlations appear to stem from individual differences and not from dialect variations. The implications of the results for rapid speaker adaptation in speech recognition will be explored. [Work supported by San Jose State University Faculty Grant.]

**1aSC16. Enhanced energy redistribution speech intelligibility algorithm with real-time implementation.** Manasa Raghavan, Mark D. Skowronski, and John G. Harris (Computational NeuroEng. Lab, Elec. and Comput. Eng., Univ. of Florida, Gainesville, FL)

Speech intelligibility enhancement is a concern for mobile platforms operating in noisy environments. Current noise-reduction techniques, such as subspace methods and spectral subtraction, operate on speech corrupted by acoustic and transmission noise. Yet preprocessing techniques, which operate on clean speech before noise corruption, have received little attention. Previously, the authors have developed the energy redistribution algorithm [J. Acoust. Soc. Am. **112**, 2305 (2002)], which, based on characteristics of clear speech as well as the Lombard effect, redistributes energy in time from voiced regions to unvoiced regions of speech. The algorithm is designed efficiently for real-time implementation, and in this work the algorithm is demonstrated on a mobile platform, TI's TMS320C6713 DSK board. Furthermore, two enhancements to the algorithm are introduced: (1) a variable unvoiced gain factor and (2) a high pass filter (HPF). The variable unvoiced gain factor adjusts the amount of energy redistributed, and the HPF, a compact algorithm shown previously to enhance clean speech in noisy environments [Niederjohn and Grotelueschen, IEEE Trans. Acoust., Speech, Signal Process. **24**(5), 277–282 (1976)], is tested on sentences from the Hearing in Noise Test (HINT) corrupted by speech-shaped noise. Results show improved speech intelligibility for both enhancements.

**1aSC17. User expertise in speech centered multimodal human computer interaction.** Rajesh Chandrasekaran, Sorin Dusan, and James L. Flanagan (CAIP Ctr., 96, Frelinghuysen Rd., Piscataway, NJ 08855)

Multimodal interfaces aim to permit natural communication by speech and gesture. Typically the speech modality bears the principal information in the interaction with gesture complementing spoken commands. A continuing challenge is how to correlate and interpret the simultaneous inputs to estimate meaning and user intent. User expertise and familiarity figure prominently in the interpretation. The present research studies the effect of

user expertise on multimodal human computer interaction. Users are classified into experienced and inexperienced depending on the amount of their exposure and interaction with multimodal systems. Each user is asked to perform simple tasks using a multimodal system. For each task the automatically recognized speech input is time stamped and the lag or lead of the gesture input is computed with respect to this time stamp. The time interval around the time stamp in which all the users' gesture inputs occur is determined. For experienced users this interval averages 56.9% less than that for inexperienced users. The implication is that for experienced users the spoken input and the corresponding gesture input are more closely related in time than for inexperienced users. This behavior can be exploited in multimodal systems to increase efficiency and reduce time of response for the system.

**1aSC18. Using Flanagan's phase vocoder to improve cochlear implant performance.** Fan-Gang Zeng (364 Med Surge II, Univ. of California, Irvine, CA 92697)

The cochlear implant has restored partial hearing to more than 100 000 deaf people worldwide, allowing the average user to talk on the telephone in quiet environment. However, significant difficulty still remains for speech recognition in noise, music perception, and tonal language understanding. This difficulty may be related to speech processing strategies in current cochlear implants that emphasized the extraction and encoding of the temporal envelope while ignoring the temporal fine structure in speech sounds. A novel strategy was developed based on Flanagan's phase vocoder [Flanagan and Golden, Bell Syst. Tech. **45**, 1493–1509 (1966)], in which frequency modulation was extracted from the temporal fine structure and then added to amplitude modulation in the current cochlear implants. Acoustic simulation results showed that amplitude and frequency modulation contributed complementarily to speech perception with amplitude modulation contributing mainly to intelligibility whereas frequency modulation contributed to speaker identification and auditory grouping. The results also showed that the novel strategy significantly improved cochlear implant performance under realistic listening situations. Overall, the present result demonstrated that Flanagan's classic work on phase vocoder still shed insight on current problems of both theoretical and practical importance. [Work supported by NIH.]

**1aSC19. Separation of formants from glottal impulses in high-resolution spectrograms using higher-order short-time Fourier transform (STFT) phase derivatives.** Sean A. Fulop (Linguist. Dept., Cal. State Univ., Fresno, CA 93740-8001)

The *channelized instantaneous frequency* of a signal  $x(T)$  is  $\text{CIF}_x(\omega, T) = (\partial/\partial T)\arg(F_h(\omega, T))$ , where  $F_h$  is the short-time Fourier transform of  $x(T)$  using window function  $h$ . The *local group delay* of a signal is  $\text{LGD}_x(\omega, T) = -(\partial/\partial\omega)\arg(F_h(\omega, T))$ . For each point  $F_h(\omega_0, T_0)$  in the STFT, the  $f$ - $t$  coordinates  $[\text{CIF}(\omega_0, T_0), t - \text{LGD}(\omega_0, T_0)]$  pinpoint the local mean of the Rihacek distribution of complex signal energy, and this reassignment of the STFT magnitude yields a high-resolution spectrogram without blurriness [Auger and Flandrin, "Improving the readability of time-frequency and time-scale representations by the reassignment method," IEEE Trans. Signal Process. **43**(5), 1068–1089 (1995)]. Nelson ["Instantaneous higher order phase derivatives," Digital Signal Process. **12**, 416–428 (2002)] demonstrated that the significant AM/FM components of a signal  $x(T)$  have  $(\partial/\partial\omega)\text{CIF}_x(\omega, T) \approx 0$ . By plotting just those STFT points meeting this condition to within a threshold, a spectrogram showing only speech resonances can be drawn using short analysis frames. By moving to a long-frame analysis, the resonance tracker becomes an effective pitch and harmonic tracker. Nelson further demonstrated that the impulses in a signal  $x(T)$  have  $(\partial/\partial T)\text{LGD}_x(\omega, T) \approx 1$ . By plotting just those points meeting this condition to within a threshold, a spectrogram of speech impulses can alternatively be drawn.

**Session 1pAA****Architectural Acoustics and Signal Processing in Acoustics: Topical Meeting on Spatial and Binaural Evaluation of Performing Arts Spaces II: Auralization and Simulation**

Mendel Kleiner, Cochair

*School of Architecture, Rensselaer Polytechnic Institute, 110 8th Street,  
Troy, New York 12180-3590*

Rendell R. Torres, Cochair

*Rensselaer Polytechnic Institute, School of Architecture, 110 8th Street, Troy, New York 12180-3590****Invited Papers*****Chair's Introduction—1:20*****Invited Papers*****1:25****1pAA1. Evaluation of room acoustic qualities and defects by use of auralization.** Jens H. Rindel (Acoust. Technol., Oersted DTU, Tech. Univ. of Denmark, Bldg. 352, DK-2800 Kgs. Lyngby, Denmark, jhr@oersted.dtu.dk)

Auralizations generated by room acoustic computer modeling programs may be used as a tool for evaluation of acoustic qualities and defects, some of which are not easily detected by objective measures. Examples include reverberance, flutter echoes, perceived room size and distance, apparent source width, listener envelopment, and sound propagation in coupled rooms. In order to reach a sufficiently high level of realism in auralizations for such room acoustic applications, it is necessary that all parts of the chain are modeled with certain accuracy. In particular, it is found that the modeling of source directivity and the late room reflections (the reverberation tail) need careful consideration in order to achieve reliable and realistic-sounding results. However, when implemented in the software the practical use can be simple and quick.

**1:45****1pAA2. Uncertainties in measurement of spatial parameters in room acoustics.** Michael Vorlaender and Ingo Witew (Inst. of Tech. Acoust., RWTH Aachen Univ., 52056 Aachen, Germany, mvo@akustik.rwth-aachen.de)

With the goal to assess spatial impression, such as apparent source width and envelopment in concert halls, lateral energy fraction is a fairly simple measure to be determined as a single number parameter. ISO 3382 describes the procedure to conduct the measurements and also defines certain quality requirements imposed on the loudspeakers and the microphones. This case study portrays how changes in the source and the microphone character are reflected in the parameter. This aspect is investigated on the grounds of measurements and simulations. In a measurement scenario, LF is calculated for different sound sources, different source setups and different microphone arrangements. In simulations, the effects of variations of the directional pattern of the microphone on the measurement result are investigated. The results suggest that despite the proven background of the lateral fraction parameter, the measurement results are distorted with noteworthy uncertainties that originate from deficiencies in measurement equipment.

**2:05****1pAA3. The value of utilizing binaural dummy head recordings in evaluating physical acoustic changes in concert halls.** Christopher Jaffe, Russell Cooper, and Carlos Rivera (Jaffe Holden Acoust., Inc., 114A Washington St., Norwalk, CT 06854)

In some instances, after a concert hall is built there may be a need to modify the physical environment of the space through the application of diffusion or absorptive surfaces, the addition of reflector systems or the repositioning of the orchestra in the space. Prior to moving forward with suggested changes to the physical environment, it has been customary to conduct evaluation rehearsals with physical mock-ups installed to confirm the acousticians recommendations. Questionnaires are given to the musicians, the conductor and the administration staff to document the effect of the changes, and physical measurements are taken before and after the installation of the mock-ups. The questionnaires can be difficult to correlate and the differences in data resulting from the physical measurements may be too small to properly evaluate. More recently, Jaffe Holden Acoustics has added dummy head recordings to the mix. These recordings are extremely representative of what a human hears and one can place these devices in various locations on stage and in the audience chamber. The recordings create a permanent record of the event and the results of subsequent A/B evaluation can be more closely correlated to render judgments.

**1pAA4. High-resolution spatial auralization techniques for room acoustics design.** Paul D. Henderson, Rendell R. Torres, and Ioana N. Pieleanu (Program in Architectural Acoust., Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180, hendep2@rpi.edu)

Historically, consultants have avoided the use of computed auralizations when evaluating critical architectural design components due to uncertainties in the physical validity and perceptual accuracy of simulations. This work demonstrates the use of advanced techniques for increasing the confidence of auralized results, applied to the acoustical design of the Experimental Media and Performing Arts Center (EMPAC), currently under construction at Rensselaer. The 1200-seat performance hall in this facility contains an acoustically transmissive overhead canopy, which creates potential difficulties for conventional simulation techniques, specifically in the modeling of early reflected and transmitted sound paths. To overcome these challenges, a new ray-tracing engine was developed which introduces a ray-spawning method in the search for deterministic sound transmission paths in the early impulse response. In addition, this method allows for surface reflection and transmission characteristics to be defined with both arbitrary frequency resolution and with dependence on incidence angle. The resulting early-part simulations are coupled with statistically computed late reverberation. These auralizations are then rendered using second-order Ambisonics over a dodecahedron loudspeaker array, recreating a full-sphere sound field for the listener. In addition, the *in situ* localization performance of human subjects using this rendering system is reviewed. [Work supported by EMPAC at Rensselaer Polytechnic Institute.]

**1pAA5. Acoustical simulation based on head-tracked auralization and measured high-resolution head related transfer function (HRTF) and impulse responses.** Wolfgang Ahnert, Stefan Feistel, and Christoph Moldrzyk (ADA Acoust. Design Ahnert, Arkonastr. 45-49, D13189 Berlin, Germany)

A very desirable feature of modern acoustical simulation programs is the easy, fast and reliable auralization of the prediction results. In this paper we consider a new auralization method, based on a head-tracked headphone system with high spatial resolution and real-time convolution. We discuss the way to measure the directional head-related transfer functions, the calculation of the directional binaural impulse responses and the realization as a real-time convolution software. Furthermore, high-resolution impulse responses have been measured to compare reality, measurement and prediction results of an example room. The measurements were performed with a newly developed software tool EASERA. We conclude that, using this new method, auralization results are obtained equivalent to the human perception in reality.

**1pAA6. Capturing the acoustic response of historical spaces for interactive music performance and recording.** Wieslaw Woszczyk and William Martens (McGill Univ., 555 Sherbrooke St. W., Montreal, QC H3A 1E3, Canada)

Performers engaged in musical recording while they are located in relatively dry recording studios generally find their musical performance facilitated when they are provided with synthetic reverberation. This well established practice is extended in the project described here to include highly realistic virtual acoustic recreation of original rooms in which Haydn taught his students to play pianoforte. The project has two primary components, the first of which is to capture for posterity the acoustic response of such historical rooms that may no longer be available or functional for performance. The project's second component is to reproduce as accurately as possible the virtual acoustic interactions between a performer and the re-created acoustic space, as performers, during their performance, move relative to their instrument and the boundaries of surrounding enclosure. In the first of two presentations on this ongoing project, the method for measurement of broadband impulse responses for these historical rooms is described. The test signal is radiated by a group of omnidirectional loudspeakers approximating the layout and the complex directional radiation pattern of the pianoforte, and the room response is sampled by a spaced microphone array. The companion presentation will describe the method employed for virtual acoustic reproduction for the performer.

**1pAA7. Virtual acoustic reproduction of historical spaces for interactive music performance and recording.** William Martens and Wieslaw Woszczyk (McGill Univ., 555 Sherbrooke St. W., Montreal, QC H3A 1E3, Canada)

For the most authentic and successful musical result, a performer engaged in recording pianoforte pieces of Haydn needs to hear the instrument as it would have sounded in historically typical room reverberation, such as that of the original room's in which Haydn taught his students to play pianoforte. After capturing the acoustic response of such historical room's, as described in the companion presentation, there remains the problem of how best to reproduce the virtual acoustical response of the room as a performer moves relative to the instrument and the rooms boundaries. This can be done with a multichannel loudspeaker array enveloping the performer, interactively presenting simulated indirect sound to generate a sense of presence in the previously captured room. The resulting interaction between live musical instrument performance and the sound of the virtual room can be captured both binaurally for the performer's subsequent evaluation, readjusted to provide the most desirable acoustic feedback to the performer, and finally remixed for distribution via conventional 5.1 channel audio media.

**1pAA8. Differences in directional sound source behavior and perception between assorted computer room models.** Michelle C. Vigeant, Lily M. Wang (Architectural Eng. Program, Univ. of Nebraska-Lincoln, Peter Kiewit Inst., 1110 S. 67th St., Omaha, NE 68182-0681, mvigeant@unlnotes.unl.edu), and Jens Holger Rindel (Tech. Univ. of Denmark, DK-2800 Kgs. Lyngby, Denmark)

Source directivity is an important input variable when using room acoustic computer modeling programs to generate auralizations. Previous research has shown that using a multichannel anechoic recording can produce a more natural sounding auralization, particularly as the number of channels is increased [J. H. Rindel, F. Otondo, and C. L. Christensen, Proceedings of the International Symposium on Room Acoustics: Design and Science 2004, Paper V01 (2004)]. Further studies evaluating the quality of auralizations using one-channel, four-channel, and 13-channel anechoic recordings have been pursued. The effect of changing the room's material properties was studied in relation to turning the source around 180 deg and on the range of acoustic parameters from the four- and 13 beams. As the room becomes increasingly diffuse, the importance of the modeled directivity decreases when considering reverberation time. However, for the three other parameters evaluated (sound-pressure level, clarity index, and lateral fraction), the changing diffusivity of the room does not diminish the importance of the directivity. The study therefore shows the importance of considering source directivity when using computer modeling. [Work supported by the National Science Foundation.]

### Contributed Papers

4:15

**1pAA9. Mapping the sound field of the Lyric Theater.** Michael Ermann, Michael Kidner, Joe McCoy, Dan Mennitt (Virginia Tech; 201 Cowgill Hall (0205), Blacksburg, VA 24061), Adam Tawney, and Bill Yoder (Virginia Tech)

Impulse response measurements were taken in each of the 235 orchestra-level house-left seats of the Lyric Theater in order to explore spatial differences in the sound field of the room. These measurements produced graphic mappings depicting gradients of sound decay time and loudness. Significant interseat variations appeared when mapping the decay times for the 63- and 125-Hz octave bands; these variations largely disappeared when mapping the decay times for octave bands greater than 250 Hz. A scaled physical model and a CATT-Acoustics software model of the space were built to explore the veracity and usefulness of modeling to determine intraroom spatial differences in sound fields.

4:30

**1pAA10. Interaural cross-correlation coefficient, apparent source width, and time/frequency characteristics of binaural running cross-correlation.** David A. Conant (McKay Conant Brook Inc., 5655 Lindero Cyn. Rd. #325, Westlake Village, CA 91362, dconant@mcbinco.com)

The interaural cross-correlation coefficient (IACC) descriptors attempt to characterize in a single number the binaural cross-correlation function at the expense of losing much detail in a manner similar to, but more problematic than, such descriptors as NC, STC, and NRC because much

frequency- and rich time-domain information is lost. Recent computer simulations by Mason illuminated these problems and further stimulated the present study. Old binaural impulse recordings run through an analog computer algorithm (drawn from physiological modeling) were re-examined in an attempt to reveal trends and relationships that the current descriptors cannot. Issues of the time and frequency variance of the measured running coefficient are discussed in terms of apparent source width (ASW) for music in Troy Music Hall and simpler spaces.

4:45

**1pAA11. Binaural simulation of virtual stage environments for evaluation by the solo violinist.** William Chu (Rensselaer Polytechnic Inst., Troy, NY 12180)

A unique computational and testing model for assessing solo violinists' response to various acoustical conditions on-stage was developed and shown to be an effective study tool. Realtime binaural auralizations of CATT-Acoustic models of several different virtual stage-acoustic designs were examined with a small group of professional violinists. Perceived differences as expressed by the subjects were discussed and explored in light of the specific measured acoustical descriptors available through the model. Unlike acoustical descriptors developed for audience receivers, ease of playing, support, and other soloist concerns do not necessarily show strong correlation between different subjects. It was found that this binaural simulation technique could provide a basis for developing a personal contextual vocabulary, to better understand the desired acoustical response of individual musicians.

## Session 1pAOa

## Acoustical Oceanography: Acoustic Characterization of the Benthic Habitat

Kenneth G. Foote, Chair

*Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution,  
Woods Hole, Massachusetts 02543*

Chair's Introduction—1:00

## Invited Paper

1:05

**1pAOa1. Geographic Information Systems (GIS) analysis of high-resolution multibeam bathymetry and remotely operated vehicle data to model rockfish habitat preference.** Erica Summers-Morris, Pat Iampietro, and Rikk Kvitek (Seafloor Mapping Lab, CSU Monterey Bay, 100 Campus Ctr., Seaside, CA 93955)

Management of declining rockfish stocks requires effective tools capable of providing accurate stock assessments of near-shore, high-relief habitat. Multibeam bathymetry, when analyzed with GIS landscape analysis tools, can create models which can identify preferred habitat based on species-specific parameters. For this study, high-resolution multibeam data of the Del Monte shale beds in Monterey Bay, California were analyzed in GIS for slope, rugosity, and relative topographic position to assess rockfish (*Sebastes spp.*) habitat preference. Video transects collected by a remotely operated vehicle (ROV) provided habitat ground-truth and fish distribution data. A series of habitat suitability models was created in GIS by combining different suitability factors from multibeam-derived grids: slope, rugosity, topographic position index (TPI), and depth. Distance to preferred categories for each parameter were determined for eight rockfish species. Of these, distance to peak features identified by TPI50 proved the most effective means of modeling fish distribution, successfully predicting an average of 80% of the eight rockfish species. Using fish distribution information, stock estimates were calculated for the study area. By combining GIS landscape analysis tools with multibeam bathymetry and ROV video data, we have created a predictive tool that can locate areas of most suitable habitat given rockfish-specific parameters.

## Contributed Papers

1:30

**1pAOa2. Quantitative benthic habitat characterization at Heceta Bank, Oregon.** Curt E. Whitmire, W. Waldo Wakefield (NOAA Fisheries Northwest Fisheries Sci. Ctr., 2725 Montlake Blvd. E, Seattle, WA 98112-2097), Robert W. Embley (NOAA Pacific Marine Environ. Lab., Newport, OR 97365), Susan G. Merle (Cooperative Inst. for Marine Resource Studies, Newport OR 97365), Brian N. Tissot, and Noelani Puniwai (Washington State Univ. Vancouver, Vancouver, WA 98686)

Recent studies have highlighted the shortcomings of regional trawl surveys for quantifying abundance of demersal fishes in rugged habitats. Many species show strong affinities to areas of high topographic relief and rugosity, therefore precluding sampling by bottom trawl gears. Consequently, advanced survey technologies such as submersibles and camera sleds have been utilized to survey demersal fishes *in situ*. One prerequisite in extrapolating densities of demersal fishes observed *in situ* is the mapping of benthic habitats beyond the observational extent of submersibles. Habitat characteristics observed from submersible and ROV dives in the late 1980s, 2000, and 2001 were used to establish habitat classes that have been statistically shown to correlate with demersal fish distributions. Those habitat characteristics were then extrapolated over the extent of a multibeam sonar survey conducted in 1998 using quantitative parameters derived from high-resolution bathymetric and acoustic backscatter data. The resultant map predicts the coverage of four lithologic habitat classes: mud (unconsolidated), sand (unconsolidated), boulder/cobble (high acoustic reflectivity), and rock outcrop (high vertical relief). Those four habitat classes will facilitate the extrapolation of fish densities to the larger spatial scales at which resident fish populations and physiographic provinces occur.

1:45

**1pAOa3. Relationships between seafloor substrates, benthic epifauna, and spatial properties of multibeam echosounder bathymetry and backscatter data.** George R. Cutter, Jr., Larry A. Mayer (UNH Ctr. for Coastal and Ocean Mapping, 24 Colovos Rd., Durham, NH 03824, gcutter@cisunix.unh.edu), and Larry G. Ward (UNH, Durham, NH 03824)

The seafloor in the mouth of the Piscataqua River Estuary, near Portsmouth, New Hampshire, contains several geomorphological regions that are evident in bathymetric data from multibeam sonar soundings. Those morphological regions generally are associated with surficial sediment type classes previously identified from sediment samples. We demonstrate that the morphologies are distinguishable using parameters from models of spatial variation for gridded depth soundings and backscatter values. Ground-truth seafloor-video transects data suggest that apparent facies and morphological regions are associated with different surficial sediment classes and rates of facies transitions. We show that the spatial variability of depth and backscatter is associated with the substrate transition rate, the number of microhabitats (microfacies), and distribution of certain epifauna identified using video. We show that, in some cases, a single region defined based upon spatial morphological attributes from depth data and apparent sedimentological attributes from backscatter data represented several benthic habitats. The results suggest that although we can use spatial variability of depth or backscatter to segment the seafloor into apparent facies or habitat class regions, some of the regions will have unpredicted associated microhabitats and organisms. The results have implications for habitat-specific assessment of organisms.

**1pAOa4. Use of a 600-kHz ADCP to characterize submerged aquatic vegetation in a very shallow estuary.** Joseph D. Warren and Bradley J. Peterson (Southampton College of Long Island Univ., 239 Montauk Hwy., Southampton, NY 11968, joe.warren@liu.edu)

Submerged aquatic vegetation (SAV) serves as a key habitat for the larval forms of many commercially important marine organisms. In coastal waters, anthropogenic factors can cause severe damage and loss to this habitat. In order to measure changes in the spatial coverage of SAV, aerial photography or diver surveys are required. We investigated an alternative approach using a high-frequency (600 kHz) acoustic Doppler current profiler to make high-resolution (in the vertical) measurements of the scattering in several very shallow bays in Long Island, NY. These data were used to determine the presence and amount of SAV in an area. Diver surveys provided ground-truthing of these data. Field measurements are also compared with an acoustic scattering model for eel grass (a common SAV type in the northeast United States). The purpose of this project was to determine whether acoustic sampling of benthic vegetation could accurately determine species type and coverage. If this is possible, then acoustic sampling could provide maps of spatial and temporal changes in SAV distribution and abundance. [Work was supported by an award from UNH CICEET.]

**1pAOa5. Acoustic scattering by benthic shells: Dominant scattering mechanisms and applications.** Timothy K. Stanton and Dezhong Chu (Woods Hole Oceanograph. Inst., Dept. of Appl. Ocean Phys. and Eng., Woods Hole, MA 02543, tstanton@whoi.edu)

When benthic shells occur in sufficiently large numbers, they can dominate acoustic backscattering by the seafloor, especially at angles of incidence away from normal. In order to use sound as a tool to remotely detect and quantify the shells, the scattering properties of the shells need to be understood, both in free-space as well as when placed on the seafloor. Through laboratory experimentation, it has been determined that the edges of certain types of shells (such as bivalves and sand dollars) can dominate the scattering over an important range of grazing angles. The surfaces of these shells and others dominate under other conditions. The dominant scattering effects are discussed in the context of interpreting acoustic backscatter data in terms of meaningful parameters such as numerical density of the shells.

**1pAOa6. Acoustic mapping of squid egg clusters and their bottom habitat in Monterey Bay, California.** Kenneth G. Foote (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543), Roger T. Hanlon (Marine Biological Lab., Woods Hole, MA 02543), Pat J. Iampietro, and Rikk G. Kvitek (California State Univ., Seaside, CA 93955)

Clusters of gelatinous egg capsules, known as mops or beds, of the market squid (*Loligo opalescens*) were mapped in a shallow-water, sandy habitat of Monterey Bay, California. The benthic egg clusters were imaged using an EdgeTech 272-TD dual-frequency sidescan sonar towed from R/V MACGINITIE, an 8-m-long survey vessel, with data recorded on a Triton Elics International Isis digital data acquisition system. Verification of target identity was accomplished independently by video photography from a remotely operated vehicle. The survey area included a 4-km stretch of sandy seafloor between Lover's Point and Cannery Row in Monterey at depths of 15–30 m. The study area had previously been mapped using the RESON SeaBat 8101 240-kHz multibeam sonar. Resulting high-resolution bathymetric data, with 1-m resolution, were used during the survey planning and execution. Squid egg clusters were clearly visible in the very-high-resolution, 400-kHz backscatter imagery, with pixel size 10–20 cm, recorded from the towed sidescan sonar. The concentration of egg clusters was greatest along a sloping feature believed to be a submarine fault. Egg mops with diameter as small as 0.5 m were distinguishable. [Support by Sea Grant is acknowledged.]

**1pAOa7. How useful is bathymetric information in the classification of high frequency sonar surveys?** Louis Atallah (Inst. of Informatics, The Br. Univ. in Dubai/ The Univ. of Edinburgh, P.O. Box 502216 Dubai, UAE) and Penny Probert Smith (Univ. of Oxford, Oxford, OX1 3PJ, UK)

In several sonar studies, bathymetric information; is used for the correction of amplitude data and the calculation of backscattering strength, which is plotted versus grazing angle and used for seabed classification. Bathymetric data is also used as an easily viewed backdrop to visualize backscattered sonar data in surveys. This work proposes an automatic method that combines amplitude features (describing backscattering strength and sonar texture) with bathymetric features (indicating seafloor variability) for sonar classification. Features are selected per window (of user defined size) and areas around grab samples in a survey are used for training. The importance of bathymetric features is investigated in this study, and highlighted by feature selection algorithms as well as by scatter plots exploring the training areas. Classification rates are significantly improved when both amplitude and bathymetry features are used. The final results show the classified windows plotted versus their exact position in the survey. The method described in this work is applied to a sidescan bathymetric sonar dataset taken in Hopvagen Bay Norway. The methods are also applicable to other sonars which provide bathymetric information; a multibeam sonar is such an example.



## Session 1pAOB

## Acoustical Oceanography: Fisheries Acoustics

Orest I. Diachok, Chair

*Poseidon Sound, 3272 Fox Mill Road, Oakton, Virginia 22124*

## Contributed Papers

3:15

**1pAOB1. Effects of fish with swim bladders on absorption and scintillation.** Orest Diachok (Poseidon Sound, Oakton, VA 22124, orestdia@aol.com)

Bioacoustic absorption spectroscopy (BAS) experiments, which were conducted in the Santa Barbara Channel in 2001 and 2002, were designed to investigate the effects of fish with swim bladders on absorption and scintillation. These experiments included a broadband source, which transmitted a sequence of 65-s-long tones between 0.25 and 10 kHz, and a vertical array which spanned most of the water column. The range was fixed. A fisheries echo sounder and trawls provided bio-acoustic parameters. Strongest absorption lines and highest values of the scintillation index were observed at night at about 1.1 kHz, the resonance frequency of 15 cm long sardines, when they were dispersed at an average depth of 13 m. Smaller absorption lines were correlated with other species. During the day sardines occupied a depth of about 50 m, where their extinction cross sections were diminished; some were dispersed and resonated at the frequency of individuals; others formed schools and resonated at collective frequencies. As a result of these phenomena, absorption lines due to sardines were much weaker, and the effect of this species on the scintillation index was not evident. [Work was supported by ONR.]

3:30

**1pAOB2. Measurements and modeling of low-frequency scattering from fish on the New Jersey shelf.** Raymond J. Soukup (Naval Res. Lab. Code 7140, Washington, DC 20375), Redwood W. Nero (Naval Res. Lab. Code 7180, Stennis Space Ctr., MS 39529), and Timothy M. Shepard (Univ. of Texas, Austin, TX 78712)

Low-frequency (400–2000 Hz) scattering strength measurements were obtained from aggregations of fish on the New Jersey shelf using a co-located transducer and vertical line array receiver. Concurrent measurements with a fisheries echosounder revealed the presence of nighttime layer(s) and daytime schools. The day–night behavior is consistent with several species of fish found on the New Jersey shelf, as are the observed resonance frequencies near 800 Hz. Data-model comparisons are used to characterize both the strength and bandwidth of the frequency response, as well as its dependence on fish type and behavior. From these data, parameters such as fish depth, density and size can be estimated, and, in turn, the scattering strength as a function of frequency and grazing angle can be predicted and then compared with those of other potential scattering mechanisms. The implications of the fish scattering to reverberation and clutter are discussed. [Work supported by ONR.]

3:45

**1pAOB3. Expected Doppler shift and spread in long range scattering from fish schools on the continental shelf.** Nicholas C. Makris, Yisan Lai, Ioannis Bertsatos, Deanne Symonds, Sunwoong Lee, and Purnima Ratilal (MIT, 77 Massachusetts Ave., Cambridge, MA 02139)

A stochastic model for the velocities of individual fish in a school, consistent with continuous wide area observations [Symonds *et al.*, J. Acoust. Soc. Am. **115**, 2618 (2004)], is developed. This is combined with

the model for scattering from a moving target submerged in a stratified ocean waveguide [Lai and Makris, J. Acoust. Soc. Am. **113**, 223–244 (2003)] to estimate the Doppler shift and spread expected in long-range scattering from fish schools on the continental shelf. The spectral characteristics expected for randomly swarming and migrating fish schools are then compared with those for stationary seafloor features and other moving targets such as underwater vehicles.

4:00

**1pAOB4. Modeling low-frequency acoustic scattering from densely populated fish schools.** Thomas R. Hahn (Rosenstiel School of Marine and Atmospheric Sci. AMP, Univ. of Miami, 4600 Rickenbacker Cswy., Miami, FL 33149)

Scattering of sound from a school of swimbladder-bearing fish is clearly related to the scattering from a cloud of air bubbles. From an “effective” medium perspective, such bubble clouds can be viewed as a single object with acoustic bulk properties determined from the physical properties of the individual bubbles and the effect of multiple scattering between them. The acoustic modeling at low frequencies of bubble clouds based on the effective medium approach has been successfully applied to bubbles in the ocean and laboratory bubble plumes by a variety of authors. Applicability and limitations of an effective medium approach to scattering of sound waves from fish schools are investigated and compared to other acoustic models. It is argued that, particularly in the regime of highly populated fish schools, where other approaches (e.g., born or random-phase approximation, etc.) fail due to strong multiple scattering effects, the effective medium model can be applied for a wide range of frequencies, fish numbers, and densities.

4:15

**1pAOB5. Diurnal trends in the mid-water biomass community of the Northwestern Hawaiian Islands observed acoustically.** Marc O. Lammers (Hawaii Inst. of Marine Biol., P.O. Box 1106, Kailua, HI 96734), Russell E. Brainard (NOAA Fisheries, Honolulu, HI 96814), and Whitlow W. L. Au (Hawaii Inst. of Marine Biol.)

The nighttime mid-water biomass occurring near six banks in the Northwestern Hawaiian Islands was investigated using 38- and 120-kHz EK60 echosounders. Locations investigated included: French Frigate Shoals, Maro Reef, Lisianski Island/Neva Shoals, Pearl and Hermes Atoll, Kure Atoll and Midway Atoll. Surveys were designed to sample transect lines parallel and normal to shore between 10 p.m. and 7 a.m. and once during daylight hours. A diurnal trend was found in the occurrence of mid-water sound scattering organisms at all six locations. Communities of organisms accumulate at night on the edges of each island between the 20- and 100-fathom isobaths. The highest densities of organisms restrict their horizontal movements to depths of 20 fathoms or deeper, but increases in biomass were also observed at shallower depths. The northern islands of Kure, Midway, and Pearl and Hermes Atolls exhibited patchier distributions than the southern islands. The composition of the biomass is presently unclear but resembles the mesopelagic boundary community found near the Main Hawaiian Islands. Simultaneous observations with the

TOAD camera system revealed clouds of zooplankton mixed with small fish and other micronekton. The nightly influx of these organisms is likely a significant, though poorly understood, component of these islands ecosystems.

4:30

**1pAOB6. Broadband acoustic backscatter and high-resolution morphology of fish: Measurement and modeling.** D. Benjamin Reeder (Naval Postgrad. School, Monterey, CA 93943), J. Michael Jech (NOAA/NMFS Northeast Fisheries Sci. Ctr., Woods Hole, MA 02543), and Timothy K. Stanton (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

Remote detection and classification of marine organisms require careful acoustic backscattering measurements and acoustic scattering models that adequately incorporate morphology. A study has been conducted on alewife (*Alosa pseudoharengus*), a swimbladder-bearing fish, which incorporated broadband (40–95 kHz) acoustic measurements, high-resolution imaging of fish morphology, and associated scattering modeling using the images. Live, adult alewife were insonified while being rotated in 1-deg increments of orientation angle in two planes of rotation (lateral and dorsal/ventral). These data provide orientation dependence of the scattering over a continuous band of frequencies, and identification of the dominant scattering features of the fish. X-ray and CT scan images of the swimbladder were digitized and incorporated into the Kirchhoff-Ray mode (KRM) model [Clay and Horne, J. Acoust. Soc. Am. **96**, 1661–1668 (1994)] and the Fourier matching method (FMM), which has recently been extended to finite-length bodies of rotation [Reeder and Stanton, J. Acoust. Soc. Am. (2004), in press]. The CT scan imagery provides rapid, noninvasive, high-resolution morphological measurements in three dimensions. The potential of the new FMM formulation is demonstrated, which is a versatile approach, valid over a wide range of shapes, all frequencies and all angles of orientation. [Work supported by ONR and NOAA.]

4:45

**1pAOB7. Optimal volume backscattering thresholds for echo integration.** Rebecca E. Thomas, Guy W. Fleischer (NOAA Fisheries, Northwest Fisheries Sci. Ctr., 2725 Montlake Blvd. E., Seattle, WA 98112, rebecca.thomas@noaa.gov), and Patrick H. Ressler (NOAA Fisheries, Newport, OR 97365)

Echo integration is a commonly used technique for assessment of fish stocks. In echo integration, a frequently used method to increase the signal-to-noise ratio is to ignore data below a set volume backscattering threshold. Although this form of thresholding is common, objective and replicable techniques for choosing this threshold are rarely used. Two opposing goals come into play when choosing an optimal threshold for estimating fish biomass using echo integration: maximizing the energy from backscatterers of interest, while simultaneously minimizing the energy from backscatterers not of interest. Both empirical and modeling techniques for choosing optimal thresholds were demonstrated using data from the 2003 integrated acoustic and trawl survey of Pacific hake. Empirical techniques were based on the receiver operating characteristic (ROC) curve. An ROC curve is a graphical plot of the number of true positives versus false positives for a binary classification system as the discrimination threshold is varied. Modeling techniques were based on

volume backscattering data generated from modeled TS and density of hake and nonhake scatterers. For the case study using the 2003 Pacific hake survey, the historical threshold used for the southern portion of the survey was shown to be nonoptimal for that year.

5:00

**1pAOB8. Identifying stable frequencies for performing animal abundance estimation with a multifrequency system.** Paul L. Roberts and Jules S. Jaffe (Marine Physical Lab, Scripps Inst. of Oceanogr., La Jolla, CA 92093-0238)

The practice of using acoustic backscatter at multiple frequencies for estimating animal abundance has become commonplace. However, the choice of frequencies to use, in order to optimize the data analysis and resulting estimates, is often not considered. The usual procedure is to use an inversion that is based on a forward model. One is generally free to choose the model used, and in some cases the frequencies used for data collection. Unfortunately, the frequencies are often chosen without regard for the stability of the inversion. The stability of a model-based multifrequency inversion depends on the condition number of the associated reflectivity matrix. This condition number in turn depends on the assumed size classes of the animals, the scattering model used, and the frequencies used to collect the data. Results indicate that the choice of frequencies has a significant effect on the condition number with greater than 3 orders of magnitude variation over a practical frequency range for the size classes considered. The analysis highlights the use of the condition number in evaluating the stability of the inversions. In addition, it provides a way to interpret estimates from existing systems, and a tool for designing future systems.

5:15

**1pAOB9. Acoustical monitoring of fish behavior in a tank.** Stephan G. Conti, Benjamin D. Maurer (Southwest Fisheries Sci. Ctr., 8604 La Jolla Shores Dr., La Jolla, CA 92037), Philippe Roux (Marine Physical Lab., La Jolla, CA 92093-0205), Christian Fauvel (Station Expérimentale d'Aquaculture IFREMER, 34250 Palavas les Flots, France), and David A. Demer (Southwest Fisheries Sci. Ctr., La Jolla, CA 92037)

In recent publications, it has been demonstrated that the total scattering cross section of fish moving in a tank can be estimated from ensembles of reverberation time series. However, the reproducibility of these measurements is influenced by parameters such as the motion or the behavior of the fish. In this work, we propose to observe acoustically the behavior of fish in a tank, and to measure their average speed. The total scattering cross section of live fish (sardines, sea bass and bocaccio) in a tank was measured repeatedly over multiple days. The species used in this study have different behaviors, which are reflected in the acoustical measurements. Depending on the behavior of the fish, such as the average displacement between two acoustic pings or the aggregation type, the total scattering cross section is different. Correlation between the acoustical measurements and the day and night behavior of the fish is demonstrated. Interpretation of such measurements can lead to monitoring acoustically and nonintrusively the behavior of fish in tanks.

1p MON. PM

## Session 1pBB

**Biomedical Ultrasound/Bioresponse to Vibration: Topical Meeting on Ultrasound Characterization  
of Cancellous and Cortical Bone II: Experimental Techniques and  
Clinical Impact and Comparison with Other Modalities**

Kendall R. Waters, Chair

*National Institute of Standards and Technology, 325 Broadway, Boulder, Colorado 80305-3328*

**Invited Papers**

**1:00**

**1pBB1. The use of ultrasound to probe structural and material properties of bone: State-of-the art.** Pascal Laugier (Universite Pierre et Marie Curie CNRS 7623, Paris, France)

The knowledge of the elastic properties of bone can be used to investigate the effects of aging, disease, and treatments. Elastic properties are also required as input for both computational and analytic models. An alternative approach to direct bone mechanical testing is to use ultrasonic testing. Furthermore, ultrasound is appropriate for multiscale analysis. *In vitro*, ultrasound has been widely used to derive nondestructively the stiffness coefficients or the elastic moduli at the whole specimen level at low frequency (structural elasticity) or at the tissue level at higher frequency (material stiffness). Quantitative ultrasound (QUS) techniques have also been developed for *in vivo* skeletal status assessment based on the assumption that ultrasonic properties reflect skeletal factors of bone fragility. *In vivo* QUS is then used to predict fracture risk. Several different techniques are currently available or under development involving different type of waves (bulk compressional waves, surface waves, guided modes). Measured ultrasonic properties in transmission or in reflection depend on a variety of material properties and macro- or micro structural characteristics. However, the complexity of the interaction mechanisms between the incident ultrasonic field and bone is still not fully elucidated and requires a better understanding.

**1:25**

**1pBB2. Use of guided ultrasonic waves for characterization of cancellous and cortical bone.** Armen Sarvazyan and Alexej Tatarinov (Artann Labs., 1753 Linvale-Harbourton Rd., Lambertville, NJ 08530)

The possibility of accessing mechanical, structural, and geometrical parameters of both the cancellous and cortical bone components by different modes of ultrasonic guided waves has been demonstrated in several laboratories in Europe and the USA. Making measurements of acoustic wave propagation parameters in a wide frequency band, e.g., from 100 kHz to several MHz, enables assessment of bone layers at different depths from the bone surface. At lower frequencies, the acoustic wave velocity is found to be sensitive to changes of the cortical thickness in middleshaft areas. In epiphyseal zones, the measured low-frequency (100-kHz) wave, velocity reflects contributions of both the spongy and compact bone components and can be sensitive to changes of the trabecular structure. At higher frequencies in the MHz range, the main acoustic mode manifested in the received signal is related to the longitudinal wave, which characterizes mainly the elastic properties of the compact bone closer to the periosteum. Measurements of the ultrasonic pulse propagation parameters in a broad frequency band using surface transmission scanning mode enables the possibility to obtain the profile of acoustic properties of long bones. [Work supported by NIH and NASA.]

**1:50**

**1pBB3. Scanning acoustic microscopy on bone—status and perspectives.** Kay Raum (Q-BAM Group, Dept. of Orthopedics, Martin Luther Univ., 06097 Halle, Germany; Laboratoire d'Imagerie Parametrique, UMR CNRS 7623, France)

High-frequency ultrasound has become one of the most powerful tools for the elastic characterization of hard materials since its invention in 1974. Many of the imaging and measurement techniques developed for NDE were also applied for the characterization of bone. However, the heterogeneous structure at several levels of organization implies multifold problems, e.g., tissue preparation, validity and applicability of the measurement techniques, resolution limitations, interpretation of results, etc. Furthermore, the technical development of commercially available high-resolution SAM lacks far behind other quantitative imaging modalities. The talk will describe the acoustic microscopes developed in our group, concepts for acoustic impedance mapping with frequencies up to 1 GHz, and techniques for determining longitudinal, shear, and surface wave velocities. The relations of acoustic to elastic parameters, either directly derived from the acoustic measurement or obtained from micromechanical tests, are presented. Combinations with other experimental and clinical techniques (US, SR-CT, nanoindentation, Raman) demonstrate the potential of SAM for an improved diagnosis, both in experimental and in clinical studies. *In vitro* results demonstrate the feasibility for assessing the elastic anisotropy and microstructural parameters of cortical bone, compositional variations within alternating osteon lamellae, structural and elastic changes in arthrotic cartilage and subchondral bone, respectively.

2:15

**1pBB4. Influence of the precision of spectral backscatter measurements on the estimation of scatterer size in cancellous bone.**

Frederic Jenson, Frederic Padilla, and Pascal Laugier (Laboratoire d'Imagerie Paramétrique, CNRS UMR 7623 Université Paris 6, 15 rue de l'Ecole de Médecine, 75006 Paris, France)

Ultrasonic backscatter measurements can be used to characterize trabecular bone structure and to estimate trabecular thickness (Tb.Th, i.e., the size of the scatterers). Our objective was to evaluate the performance of this estimator as well as others spectral estimators, like the frequency dependence and the mid-band amplitude of the backscatter coefficient. The performance of these estimators is degraded mainly by two factors: interference noise due to random positioning of the scatterers and attenuation. We have simulated rf-lines backscattered from trabecular bone assuming a random positioning of the trabeculae (leading to a fully developed speckle), a Gaussian form factor for the scatterers and a linear-frequency dependent attenuation. It is found that the variance in the estimation of the frequency dependence of the backscatter coefficient is as high as the variance due to the biological variability in Tb.Th, in agreement with the results showed by K. Wear [J. Acoust. Soc. Am. SA, **110** (2001)]. In contrast, the variances on the mid-band amplitude and on the estimated trabecular thickness are lower than the variance due to the biological variability. We also show that the effect of attenuation may be compensated by using an appropriate attenuation-compensation function. These results suggest that the inverse problem can be appropriately addressed.

2:30

**1pBB5. Assessment of cortical bone microstructure and material properties using high resolution scanning acoustic microscopy.**

Kay Raum (Q-BAM Group, Dept. of Orthopedics, Martin Luther Univ., 06097 Halle, Germany, kay.raum@medizin.uni-halle.de), Florent Chandelier, Ingrid Leguerney, Maryline Talmant, Amena Saied, Pascal Laugier (Laboratoire d'Imagerie Paramétrique, 75006 Paris, France), and Françoise Peyrin (CREATIS and European Synchrotron Radiation Facility, 38043 Grenoble, France)

Combined evaluation of bone microstructural and mechanical information remains a challenging task which is required for bone phenotyping or accurate finite element modeling. Our objective was to assess the value of quantitative scanning acoustic microscopy (SAM) for bone characteriza-

tion in comparison to synchrotron radiation computed tomography (SR-CT). Ten specimens of human cortical bone (radius) were investigated using SR-CT and SAM (200 MHz) with spatial resolution of 10 and 8  $\mu\text{m}$ , respectively. An image fusion and analysis software was developed to derive site-matched estimates of (1) microstructural parameters, e.g., Haversian cavity density and mean diameter and porosity, and (2) tissue properties such as mineral density (MD, SR-CT) and acoustic impedance (Z, SAM) for distinct anatomical regions of interest (osteons, interstitial tissue). Local stiffness  $c_{33}$  was derived from the combination of MD and Z. An almost perfect correlation was found for all microstructural indices derived by both techniques. Impedance was correlated to the square of MD ( $R^2=0.39$ ,  $p<1e-4$ ). The derived stiffness  $c_{33}$  ( $35.9\pm12.8$ ) was highly correlated with Z ( $R^2=0.99$ ,  $p<1e-4$ ). These findings suggest that SAM fulfills the requirement for a simultaneous evaluation of cortical bone microstructure and material properties at the tissue level.

2:45

**1pBB6. Correction for soft tissue in cortical bone assessment by ultrasound.**

E. Bossy, M. Talmant, P. Laugier (Laboratoire d'Imagerie Paramétrique, UMR 7623, 15 rue de l'école de Médecine, 75006 Paris, France), C. Roux, S. Kolta (CEMO, 75014 Paris, France), and D. Haguénauer (Hôpital Ste Perrine, 75016 Paris, France)

One of the key points in ultrasound measurements on cortical bone is the correction for soft tissue. We designed a new probe based on bi-directional axial transmission which automatically compensates velocity measurements for the soft tissue effect without preliminary evaluation of soft tissue properties. The probe consists in a linear arrangement of transducers with two sources placed on both sides of a unique group of receivers. The velocity of waves propagating parallel to the bone axis is deduced from a combination of the time delays derived from waves propagating in opposite directions at successive receivers separated by a known distance. This technique efficiently corrects for the major source of error on velocity encountered in clinical measurements which is caused by the variation of soft tissue thickness along the probe. The bi-directional technique was validated on test samples for which the residual precision error on velocity measurements was reduced to 0.2%. *In vivo* measurements yielded a value of 0.5% for the interoperator reproducibility. The clinical range of variation of the velocity measured by bi-directional technique is evaluated using clinical measurements on more than 200 subjects. Bi-directional transmission is a promising technique to minimize the variability of *in vivo* velocity measurements.

3:00–3:15 Break

**Invited Papers**

3:15

**1pBB7. Ultrasound critical angle-reflectometry: Measuring ultrasound velocities in the clinic and in the laboratory.**

Peter P. Antich, Matthew A. Lewis, Edmond Richer, Billy J. Smith (Adv. Radiological Sci., Univ. of Texas Southwestern Med. Ctr., Dallas, TX 75390-9058, peter.antich@utsouthwestern.edu), and Charles Y. C. Pak (Univ. of Texas Southwestern Med. Ctr., Dallas, TX 75390)

Ultrasound critical angle-reflectometry is a modality developed and tested for measuring US velocity in the laboratory or the clinic. The applicator developed for use in the clinic consists of a transmitter concentric to a receiver, immersed in water and capable of analyzing multiple reflections from soft and mineralized tissues over angles from  $-45^\circ$  to  $45^\circ$ . If a critical angle  $\theta_c$  is detected in the spectrum, the velocity in bone is simply  $V=c/\sin\theta_c$ . The analysis can be repeated at different orientations; considering only p-wave velocities, as bone has transverse symmetry, there is a unique relationship between velocity and coefficients of stiffness:  $\rho V^2(\phi) = C_{11} \cos^4(\phi) + C_{33} \sin^4(\phi) + 2 \cos^2(\phi) \sin^2(\phi) (C_{13} + 2C_{44})$ . It can be shown that  $C_{11}$  and  $C_{33}$  give two orthogonal moduli of elasticity. Experiments show that the modulus of elasticity so derived is equal to that measured by mechanical testing. In addition to these angles we can measure the arrival time of the reflected signal and identify the depth from which the signal originates; for cortical layers of sufficiently small thickness it is possible to identify cancellous bone critical angles. Examples are provided which show that these quantities vary in response to metabolic and physical stimuli.

**1pBB8. Assessment of trabecular bone quality in human cadaver calcaneus using scanning confocal ultrasound and dual x-ray absorptiometry (DEXA) measurements.** Yixian Qin (Dept. of Biomed. Eng., SUNY Stony Brook, 350 Psych-A Bldg., Stony Brook, NY 11794, yi-xian.qin@sunysb.edu), Yi Xia, Wei Lin, Clinton Rubin, and Barry Gruber (SUNY Stony Brook, Stony Brook, NY 11794)

Microgravity and aging induced bone loss is a critical skeleton complication, occurring particularly in the weight-supporting skeleton, which leads to osteoporosis and fracture. Advances in quantitative ultrasound (QUS) provide a unique method for evaluating bone strength and density. Using a newly developed scanning confocal acoustic diagnostic (SCAD) system, QUS assessment for bone quality in the real body region was evaluated. A total of 19 human cadaver calcanei, age 66 to 97 years old, were tested by both SCAD and nonscan mode. The scanning region covered an approximate  $40 \times 40 \text{ mm}^2$  with 0.5 mm resolution. Broadband ultrasound attenuation (BUA, dB/MHz), energy attenuation (ATT, dB), and ultrasound velocity (UV, m/s) were measured. The QUS properties were then correlated to the bone mineral density (BMD) measured by DEXA. Correlations between BMD and QUS parameters were significantly improved by using SCAD as compared to nonscan mode, yielding correlations between BMD and SCAD QUS parameters as  $R=0.82$  (BUA), and  $R=0.86$  (est. BMD). It is suggested that SCAD is feasible for *in vivo* bone quality mapping. It can be potentially used for monitoring instant changes of bone strength and density. [Work supported by the National Space Biomedical Research Institute (TD00207), and New York Center for Biotechnology.]

**1pBB9. Ultrasonic assessment of bone: Comparison with other characterization modalities.** Patrick H. Nicholson (Dept. of Health Sci., Univ. of Jyväskylä, P.O. Box 35, 40014, Finland, patrick\_nicholson@btopenworld.com)

Compared to other techniques, quantitative ultrasound of bone (QUS) has a unique potential arising from the mechanical nature of the wave phenomena involved, the ability to use wavelengths spanning the dimensions of key structural features of bone, and the absence of ionizing radiation. However, QUS suffers from fundamental problems of interpretation when compared to most other bone assessment modalities. For example, dual energy x-ray absorptiometry gives a direct estimate of a well-understood physical parameter: bone mineral density. In contrast, QUS yields acoustic properties that cannot be so easily interpreted, and whose clinical value is derived largely from empirical relationships with physical properties of interest, such as bone density, or with other factors such as risk of fracture. Hence for QUS we cannot trace a clear causal path connecting an ultrasonic measurement to specific bone properties and on, in turn, to clinical outcomes and decisions. Without developing the theoretical framework required to achieve this, QUS will not achieve its full potential and is likely to remain sidelined and mistrusted in comparison to the x-ray-based modalities which are its principal competitors.

### Contributed Papers

**1pBB10. In vitro comparative study of three devices based on ultrasonic axial transmission.** M. Muller, M. Talmant, P. Laugier (Laboratoire d'Imagerie Paramétrique, UMR CNRS 7623, 15 rue de l'école de Médecine, 75006 Paris, France), P. Moilanen, V. Kilappa, J. Timonen, P. Nicholson, and S. Cheng (Univ. of Jyväskylä, Jyväskylä, Finland)

In the axial transmission technique the velocity of waves propagating in the direction of bone axis is used to characterize cortical bone (radius, tibia, etc.). Corresponding clinically used devices are based on a long wavelength approach (typical frequency around 200 kHz) or shorter wavelength (typical frequency around 1 MHz) compared to bone thickness. They differ also by the methods of wave velocity evaluation. The aim of this study is to compare the sensitivity to bone properties of three representative devices using an *in vitro* investigation of the same specimens coupled with x-ray determination of bone properties. The moderate correlation between velocities suggests an important impact of site-matching. Among higher frequency devices, the one associated to the bidirectional probe provided generally higher correlation with bone properties than conventional axial transmission. The high-frequency devices are less sensitive to cortical thickness, CSA and trabecular BMD than the low-frequency device because high-frequency waves interrogate a thinner cortical layer than low-frequency waves. Our results suggest that different axial transmission approaches reflect different bone properties. Therefore, a multi-frequency technique might be useful in probing different bone properties at the same time (e.g., cortical thickness and BMD).

**1pBB11. A new system for clinical ultrasound assessment of bone.** Jonathan J. Kaufman, Gangming Luo (CyberLogic, Inc., 611 Broadway, Ste. 707, New York, NY 10012), Miriam Englander, and Robert S. Siffert (Mount Sinai School of Medicine, New York, NY 10029)

A new ultrasound device for noninvasive assessment of bone known as the QRT 2000 for Quantitative Real-Time—that is entirely self-contained, portable, and handheld is described. The QRT 2000 is powered by 4 AA rechargeable batteries and permits near real-time evaluation of a novel set of ultrasound parameters and their on-line display to the user. A clinical study has just been completed with the QRT 2000 in which 60 female subjects ranging in age from 25 to 88 were ultrasonically interrogated at their heels. The same heel was measured also using DEXA (PIXI, GE Medical Systems) and the bone mineral content (BMC) was compared with one ultrasound parameter which has been found to be extremely sensitive to bone mass. The parameter, known as the net time delay (NTD), and BMC had an associated *R*-squared value of 0.73, about a 13% improvement over presently marketed devices. This, coupled with the lower cost and portability of the system, makes the QRT 2000 ideally suited for use by primary care physicians in this country and abroad, and including for use in the developing world. Further improvements are being pursued through array methods (to improve reproducibility and correlations with BMD) and by incorporating other parameters particularly sensitive to architectural structure. [This research was supported by SBIR Grant No. 2R44AR045150 from the National Institute of Arthritis and Musculoskeletal and Skin Diseases of the NIH.]

### Panel Discussion

## Session 1pPA

## Physical Acoustics: Sound in Ground and Related Topics

Steven R. Baker, Chair

Naval Postgraduate School, Code PH/BA, Monterey, California 93943

## Contributed Papers

2:00

**1pPA1. Maximum effective aperture size for source direction estimates in a complex terrain.** Mark Moran, D. Keith Wilson, and Roy Greenfield (USACE ERDC-CRREL, 72 Lyme Rd., Hanover, NH 03755)

Source resolution and clutter suppression from ground sensor arrays scales directly with sensor array aperture. However, boundary layer complexities can disrupt acoustic and seismic wavefronts at a variety of spatial scales. Hence, usage of an array dimension that is larger than these wavefront distortions can lead to degraded resolution and poor clutter suppression. Using experimental data, we demonstrate the maximum effective aperture for determining a line of bearing to a light moving ground vehicle. Our data were collected in mountainous terrain with open fields, dense forests, and steep hills. We deployed a field of 36 seismic sensors (geophones) and 36 high quality microphones. These sensors were distributed in triangular subclusters within a larger equilateral triangle that was roughly 100 m on each side. Our microphone separations were as small as 10 cm, with common separations of 1, 5, 10, and 25 m. Our seismic subclusters had common separations of approximately 5, 15, and 30 m. We apply a high resolution 2-D maximum-likelihood (ML) wavenumber estimation method to analyze array gain characteristics at multiple aperture sizes. Our results show large angular and range dependencies that correlate strongly with terrain complexity and disruptions of line-of-sight with a moving vehicle.

2:15

**1pPA2. Theoretical analysis of a directional source for four dimensional seismic monitoring.** Madjid Berraki, Bertrand Dubus (IEMN, Dept. ISEN, UMR CNRS 8520, 41 Blvd. Vauban 59046 Lille, Cedex France), and Axelle Baroni (Institut Français du Pétrole, 92852 Reuil Malmaison, Cedex France)

The aim of four dimensional seismic monitoring is to monitor the changes in the subsurface in time; to achieve this, measurements are continuously repeated with the same space configuration. Reliability and repeatability of the acquisition footprint are thus key requirements for four dimensional surveys. The weathered zone (WZ) is the main cause of source signal fluctuations, polluting the signal coming from reservoir variations. This zone is directly located beneath the surface and its mechanical properties vary with weather. To eliminate this perturbation, a directional source, buried under the WZ and radiating downwards, is analyzed here. Assuming that the source size to wavelength ratio is small, a model is developed for the directional source buried in a layered half space: the Green's tensor in a layer is calculated using the discrete wavenumber method [M. Bouchon, Bull. Seismol. Soc. Am. **71**(4), 959-971 (1981)] and the solution is then propagated using the reflectivity method [J. Muller, Geophysics **58**, 153-174 (1985)]. Using a typical seismic signal called Ricker, synthetic seismograms are computed for the directional source buried in a half space covered by a layer (the WZ). The obtained results validate the directional source concept. [Work supported by an I.F.P. grant.]

2:30

**1pPA3. Coupling finite element and iterative methods on a solid-solid boundary—Computation of the radiation of buried directional piezoelectric transducers for seismic applications.** Madjid Berraki, Bertrand Dubus (IEMN, ISEN Dept., UMR CNRS 8520, 41 Blvd. Vauban, 59046 Lille Cedex, France), and Axelle Baroni (Institut Français du Pétrole, Malmaison Cedex, France)

Based on a theoretical directional source concept for four dimensional seismic monitoring (described in another communication), this work is dedicated to the design of the directional piezoelectric transducer. To describe the interaction of a piezoelectric source with its environment, a model coupling the finite element method with the impedance matrix of the soil is developed. The piezoelectric source and a surrounding spherical soil volume are described using finite element; on the surface of the spherical volume, a condition taking into account the external medium is applied by using an iterative method [Thiruvankatachar *et al.*, Proc. R. Soc. London, Ser A **309**, 331-344 (1969)]. This numerical process is then implemented in the finite element code ATILA which is devoted to the modeling of piezoelectric structures. A transducer acting as the directional concept is designed and its far field radiation patterns are computed. [Work supported by an I.F.P. grant.]

2:45

**1pPA4. Time-domain calculations of acoustic interactions with rigid porous surfaces.** D. Keith Wilson (U.S. Army Eng. Res. Dev. Ctr., 72 Lyme Rd., Hanover, NH 03755), David F. Aldridge, Neill P. Symons (Sandia Natl. Labs., Albuquerque, NM 87185), Vladimir E. Ostashev (NOAA/Environ. Tech. Lab., Boulder, CO 80305), Sandra L. Collier, and David H. Marlin (U.S. Army Res. Lab., Adelphi, MD 20783)

In previous work, a set of time-domain equations for sound propagation in a rigid porous medium, including viscous and thermal dissipation effects, was derived [Ostashev *et al.*, J. Acoust. Soc. Am. **115**, 2624 (2004)]. From those equations, time-domain counterparts to frequency-domain impedance boundary conditions (BCs) were also derived. In this paper, several approaches to computational implementation of time-domain acoustic interactions with rigid porous surfaces are discussed that are based on these equations. Most of the approaches involve convolutions between relaxation or transformed impedance functions and the acoustic wavefield variables. For these approaches, when the relaxation times for viscous and thermal diffusion in the pores (which are inversely proportional to the static flow resistivity) are large, the convolution integrals must be evaluated over many periods of the acoustic disturbance and therefore become very computationally demanding. Some example finite-difference time-domain (FDTD) calculations show the significantly increased attenuation within the porous medium resulting from the convolution terms. Computational instabilities are observed when the wave is strongly attenuated over the FDTD spatial grid interval. Alternative ap-

proaches, based on Pade and series approximations to the time-domain impedance BC, are also considered that provide more compact numerical solutions.

3:00

**1pPA5. The role of hysteresis in propagating acoustic wave in porous materials.** Zhiqun Lu and James Sabatier (Natl. Ctr. for Phys. Acoust., The Univ. of Mississippi, Coliseum Dr., Univ., MS 38677)

Hysteresis and end-point memory are universal behaviors of porous materials including rock, sandstone, soil, sediment, ceramic, and some cracked materials. Recently, there have been growing interests in studying hysteresis and linking this static property to dynamic behaviors, particularly in modeling nonlinear acoustic phenomena in porous materials. This study aims at exploring the role of hysteresis in propagation of an acoustic wave in soils. A modified triaxial cell was used to measure stress/strain and sound speed/strain relationships in a triaxial test. Soil samples were subjected to a load stress path that was designed to undergo a series of meso-scopic unload-reload stress cycles with axial strain in the range of  $10\text{Exp}(-5) - 10\text{Exp}(-4)$ . It is found that the slope of the hysteresis loop is linearly correlated with the dynamic elastic modulus, which implies that the fluctuation of acoustic pressure may create micro-hysteresis loops that make the dynamic elastic modulus higher than static elastic modulus. This discovery may lead to a correction in the classical equation of state by adding a hysteresis term. [Work supported by the U.S. Department of Agriculture.]

3:15

**1pPA6. Impedance and Brewster angle measurement for thick porous layers.** Del Leary, Craig J. Hickey (The Univ. of Mississippi, Coliseum Dr., Univ., MS 38677), Jean F. Allard, and Michel Henry (l'Université du Maine, 72085 Le Mans Cedex 9, France.)

For thin nonlocally reacting porous layers, a method derived from the work of Chien and Soroka [J. Sound Vib. **43**, 9–20 (1975)] has been used to localize the pole of the reflection coefficient located at an angle close to grazing incidence and to measure the surface impedance at this angle. A very simple experimental setup is used to obtain measurements on materials with large flow resistivity, at frequencies larger than 1 kHz, using samples with areas on the order of  $1\text{ m}^2$ . This method is used in the present work to measure the surface impedance of acoustically thick porous layers of Ottawa sand and glass beads. There is good agreement between the measurements and predicted values. The method is also applied to study the effects of surface sealing. Sealing is modeled as a thin screen of higher flow resistivity at the surface. Reasonable agreement between the measured and predicted additional flow resistance is obtained.

3:30–3:45 Break

3:45

**1pPA7. Modal analysis of broadband acoustic signals propagating in the top layer of the ground.** Vladimir N. Fokin, James M. Sabatier, and Wheeler B. Howard (Natl. Ctr. for Phys. Acoust., 1 Coliseum Dr., University, MS 38655, mfok@olemiss.edu)

Experimental measurements of the frequency dependence of sound propagation losses in the top layers of the ground were performed in the frequency range between 80 and 420 Hz. These measurements revealed that the distribution of energy on the frequency-range plane has a regular structure. The features of this structure were modeled in the modal approach for a gradient model of the ground. It was shown that average sound speed in the ground and the approximate depth of the layers may be estimated through analysis of the experimentally measured frequency-range distribution of energy. The technique explores the waveguide properties of the top layer of the ground. According to the mode theory, two asymptotes coming from the point (0,0) should exist on the wave number-frequency plane. From the slope of these asymptotes, the minimum and

maximal sound speeds in the waveguide were found to be 100 and 530 m/s. The approximate depth of the layers was estimated. Comparisons of obtained sound speeds and depth of layers with independently estimated values show satisfactory agreement. [Work supported by ONR Grant N00014-02-1-0878.]

4:00

**1pPA8. Propagation of interface waves in a clay soil.** Thomas G. Muir (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677), Richard Burgett (Planning Systems Inc., University, MS 38677), and Robert M. O'Neill (LD Consulting, Oxford, MS 38655)

Measurements and analyses on the propagation of interface waves, previously conducted on a sandy beach in Monterey, CA, are extended to a clay soil in Oxford, MS, in support of research on seismic sonar. Data are presented on attenuation versus lateral range and frequency, which may result from attenuation and scattering, as well as refraction and modal effects. Data are also presented on spatial coherence and group velocity, as a function of angle, along semi-circular arcs, also as a function of lateral range. The interface waves were generated in pulses, with both moving coil and moving mass sources (shakers), and were measured with three axis seismometers, at frequencies ranging from a few tens of Hz to several hundred Hz, at lateral ranges from a few meters to a few tens of meters. The sources excited numerous seismo-acoustic wave types; but the analysis focused on (1) Rayleigh waves having radial and vertical components in the vertical plane, and (2) transverse waves in the horizontal plane, thought to be Love waves. The data are interpreted with respect to environmental "ground truth" measurements. [Work supported by the U.S. Marine Corps Systems Command.]

4:15

**1pPA9. Variation in acoustic to seismic signatures associated with natural depth variability of hardpans.** Wheeler B. Howard, Craig J. Hickey, James M. Sabatier (Natl. Ctr. for Phys. Acoust./ Univ. of Mississippi, University, MS 38677), and David A. DiCarlo (Natl. Sedimentation Lab./ ARS)

Acoustic to seismic coupling techniques have been successfully used in the investigation of near-surface soils. Spatial variation in the soil profile and associated near-surface properties are manifested as differences in the acoustic to seismic transfer function. Two field sites were chosen which differ by the average depth of a naturally occurring hardpan. In this presentation, the measured acoustic to seismic response at multiple point locations within these two agricultural field sites is presented. Variability in attributes of the acoustic to seismic transfer function, namely the frequency, half-width, and amplitude of the resonance, are compared to variability of the depths to the hardpan.

4:30

**1pPA10. Acoustic-to-seismic transfer function attribute for discrimination of false alarms and landmine detection.** Margarita S. Fokina and James M. Sabatier (Natl. Ctr. for Phys. Acoust., 1 Coliseum Dr., University, MS 38655, vfok@olemiss.edu)

Acoustic detection of landmines based on the analysis of both spatial and frequency dependencies of the acoustic-to-seismic transfer function (A/S TF) utilizes the difference between mine impedance and the impedance of the surrounding ground. However, some deeply buried mines and some types of mines are hard to detect due to the natural variability of the ground. This work addresses the problem of false alarms and clutter (high values of the A/S TF in some frequency bands) that mimic the physics of a buried landmine. A time-scale, linear method (wavelet analysis) was used for improving the probability of landmines detection. Wavelet analysis of the calculated and measured signals permit one to find stable characteristics typical of the undisturbed ground, the disturbed ground, and the ground with a mine. The wavelet analysis showed that these characteristics may be used for the discrimination of landmines from false alarms. These characteristics also may be used as the additional criterion to find mines

which are hard to locate by traditional methods. The advantages of the suggested technique are illustrated using the experimental data. Possibilities for buried landmine classification are discussed. [Work supported by ONR Grant N00014-02-1-0878.]

4:45

**1pPA11. Experiments on nonlinear acoustic landmine detection: Tuning curve studies of soil-mine and soil-mass oscillators.** Murray S. Korman (Phys. Dept., U.S. Naval Acad., Annapolis, MD 21402), Thomas R. Witten, and Douglas J. Fenneman (U.S. Army Night Vision and Electron. Sensors Directorate, Fort Belvoir, VA 22060)

Donskoy [SPIE Proc. **3392**, 211–217 (1998); **3710**, 239–246 (1999)] has suggested a nonlinear technique that is insensitive to relatively non-compliant targets that can detect an acoustically compliant buried mine. Airborne sound at two primary frequencies eventually causes interactions with the soil and mine generating combination frequencies that can affect the vibration velocity at the surface. In current experiments,  $f_1$  and  $f_2$  are closely spaced near a mine resonance and a laser Doppler vibrometer profiles the surface. In profiling, certain combination frequencies have a much greater contrast ratio than the linear profiles at  $f_1$  and  $f_2$ —but off the mine some nonlinearity exists. Near resonance, the bending (a softening) of a family of tuning curves (over the mine) exhibits a linear relationship between peak velocity and corresponding frequency, which is characteristic of nonlinear mesoscopic elasticity effects that are observed in geomaterials like rocks or granular media. Results are presented for inert plastic VS 1.6, VS 2.2 and M14 mines buried 3.6 cm in loose soil. Tuning curves for a rigid mass plate resting on a soil layer exhibit similar results, suggesting that nonresonant conditions off the mine are desirable. [Work supported by U.S. Army RDECOM, CERDEC, NVESD, Fort Belvoir, VA.]

5:00

**1pPA12. Nonlinear acoustic experiments involving landmine detection: Connections with mesoscopic elasticity and slow dynamics in geomaterials.** Murray S. Korman (Phys. Dept., U.S. Naval Acad., Annapolis, MD 21402), Douglas J. Fenneman (U.S. Army Night Vision and Electron. Sensors Directorate, Fort Belvoir, VA 22060), and James M. Sabatier (Univ. of Mississippi, Univ., MS 38677)

The vibration interaction between the top-plate of a buried VS 1.6 plastic, anti-tank landmine and the soil above it appears to exhibit similar characteristics to the nonlinear mesoscopic/nanoscale effects that are observed in geomaterials like rocks or granular materials. In nonlinear detection schemes, airborne sound at two primary frequencies  $f_1$  and  $f_2$  (chosen several Hz apart on either side of resonance) undergo acoustic-to-seismic coupling. Interactions with the compliant mine and soil generate combination frequencies that, through scattering, can effect the vibration velocity at the surface. Profiles at  $f_1$ ,  $f_2$ ,  $f_1 - (f_2 - f_1)$  and  $f_2 + (f_2 - f_1)$  exhibit a single peak while profiles at  $2f_1 - (f_2 - f_1)$ ,  $f_1 + f_2$  and  $2f_2 + (f_2 - f_1)$  are attributed to higher order mode shapes. Near resonance ( $\sim 125$  Hz for a mine buried 3.6 cm deep), the bending (softening) of a family of increasing amplitude tuning curves (involving the surface vibration over the landmine) exhibits a linear relationship between the peak particle velocity and corresponding frequency. Subsequent decreasing amplitude tuning curves exhibit hysteresis effects. Slow dynamics explains the amplitude difference in tuning curves for first sweeping upward and then downward through resonance, provided the soil modulus drops after periods of high strain. [Work supported by U.S. Army RDECOM, CERDEC, NVESD, Fort Belvoir, VA.]

1p MON. PM



**Session 1pSC****Speech Communication and Signal Processing in Acoustics: Fifty Years of Progress in Speech Communication: Honoring the Contributions of James L. Flanagan**

Sorin Dusan, Cochair

*Rutgers University, CAIP Center, 96 Frelinghuysen Road, Piscataway, New Jersey 08854-8088*

Lawrence Rabiner, Cochair

*Rutgers University, 96 Frelinghuysen Road, Piscataway, New Jersey 08854***Chair's Introduction—1:30*****Invited Papers*****1:35****1pSC1. Fifty years of progress in acoustic phonetics.** Kenneth N. Stevens (Res. Lab of Elec. and Dept. Elec. Eng., MIT, 50 Vassar St., Rm. 36-517, Cambridge, MA 02139-4307)

Three events that occurred 50 or 60 years ago shaped the study of acoustic phonetics, and in the following few decades these events influenced research and applications in speech disorders, speech development, speech synthesis, speech recognition, and other subareas in speech communication. These events were: (1) the source-filter theory of speech production (Chiba and Kajiyama; Fant); (2) the development of the sound spectrograph and its interpretation (Potter, Kopp, and Green; Joos); and (3) the birth of research that related distinctive features to acoustic patterns (Jakobson, Fant, and Halle). Following these events there has been systematic exploration of the articulatory, acoustic, and perceptual bases of phonological categories, and some quantification of the sources of variability in the transformation of this phonological representation of speech into its acoustic manifestations. This effort has been enhanced by studies of how children acquire language in spite of this variability and by research on speech disorders. Gaps in our knowledge of this inherent variability in speech have limited the directions of applications such as synthesis and recognition of speech, and have led to the implementation of data-driven techniques rather than theoretical principles. Some examples of advances in our knowledge, and limitations of this knowledge, are reviewed.

**1:50****1pSC2. Fifty years of progress in speech waveform coding.** Bishnu S. Atal (Dept. of Elec. Eng., Univ. of Washington, Seattle, WA 98195, bsatal@bishnu.net)

Over the past 50 years, sustained research in speech coding has made it possible to encode speech with high speech quality at rates as low as 4 kb/s. The technology is now used in many applications, such as digital cellular phones, personal computers, and packet telephony. The early research in speech coding was aimed at reproducing speech spectra using a small number of slowly varying parameters. The focus of research shifted later to accurate reproduction of speech waveforms at low bit rates. The introduction of linear predictive coding (LPC) led to the development of new algorithms, such as adaptive predictive coding, multipulse and code-excited LPC. Code-excited LPC has become the method of choice for low bit rate speech coding and is used in most voice transmission standards. Digital speech communication is rapidly moving away from traditional circuit-switched to packet-switched networks based on IP protocols (VoIP). The focus of speech coding research is now on providing to low cost, reliable, and secure transmission of high-quality speech on IP networks.

**2:05****1pSC3. Fifty years of progress in speech analysis (LPC).** Fumitada Itakura (Meijo Univ., Tempaku-ku, Nagoya 468-8502, Japan)

Although the history of statistical linear prediction is very long, it was first applied to speech analysis in 1966 at NTT to estimate the all-pole speech spectrum envelope in order to implement ASR and vocoder. All-pole spectral parameters are investigated in detail to find a better representation with respect to quantization and interpolation characteristics, partly at NTT and Bell Labs, leading to PARCOR and LSP. These parameters were applied to narrow band speech coders or LPC vocoders and speech synthesizer chips in the late 1970s and early 1980s, but the speech quality was insufficient for digital mobile telephone application. The problem was later solved by using hybrid CELP and MLP coding mainly at BTL. LPC analysis was also applied to acoustic front-end for ASR. Again it was found that LPC suffers with additive noise and linear/nonlinear distortions. Whereas computational efficiency of LPC was used to be the most prominent advantage, today we have gained thousands of times the processing power at a ten thousandth of the cost in 40 years. It is hoped a novel series of speech analysis methods, whose competence is comparable to human auditory system, should be developed at any expense of computational complexity.

**1pSC4. Fifty years of progress in speech coding standards.** Richard Cox (AT&T Labs, 180 Park Ave., Florham Park, NJ 07932, rvc@research.att.com)

Over the past 50 years, speech coding has taken root worldwide. Early applications were for the military and transmission for telephone networks. The military gave equal priority to intelligibility and low bit rate. The telephone network gave priority to high quality and low delay. These illustrate three of the four areas in which requirements must be set for any speech coder application: bit rate, quality, delay, and complexity. While the military could afford relatively expensive terminal equipment for secure communications, the telephone network needed low cost for massive deployment in switches and transmission equipment worldwide. Today speech coders are at the heart of the wireless phones and telephone answering systems we use every day. In addition to the technology and technical invention that has occurred, standards make it possible for all these different systems to interoperate. The primary areas of standardization are the public switched telephone network, wireless telephony, and secure telephony for government and military applications. With the advent of IP telephony there are additional standardization efforts and challenges. In this talk the progress in all areas is reviewed as well as a reflection on Jim Flanagan's impact on this field during the past half century.

**1pSC5. Fifty years of progress in speech synthesis.** Juergen Schroeter (AT&T Labs-Res., Rm. D163, 180 Park Ave., Florham Park, NJ 07932)

A common opinion is that progress in speech synthesis should be easier to discern than in other areas of speech communication: you just have to listen to the speech! Unfortunately, things are more complicated. It can be said, however, that early speech synthesis efforts were primarily concerned with providing intelligible speech, while, more recently, "naturalness" has been the focus. The field had its "electronic" roots in Homer Dudley's 1939 "Voder," and it advanced in the 1950s and 1960s through progress in a number of labs including JSRU in England, Haskins Labs in the U.S., and Fant's Lab in Sweden. In the 1970s and 1980s significant progress came from efforts at Bell Labs (under Jim Flanagan's leadership) and at MIT (where Dennis Klatt created one of the first commercially viable systems). Finally, over the past 15 years, the methods of unit-selection synthesis were devised, primarily at ATR in Japan, and were advanced by work at AT&T Labs, Univ. of Edinburgh, and ATR. Today, TTS systems are able to "convince some of the listeners some of the time" that synthetic speech is as natural as live recordings. Ongoing efforts aim at replacing "some" with "most" for a wide range of real-world applications.

**1pSC6. Fifty years of progress in speaker verification.** Aaron E. Rosenberg (Ctr. for Adv. Information Processing, Rutgers Univ., CoRE Bldg, 96 Frelinghuysen Rd., Piscataway, NJ 08854-8088, aer@caip.rutgers.edu)

The modern era in speaker recognition started about 50 years ago at Bell Laboratories with the controversial invention of the voiceprint technique for speaker identification based on expert analysis of speech spectrograms. Early speaker recognition research concentrated on finding acoustic-phonetic features effective in discriminating speakers. The first truly automatic text dependent speaker verification systems were based on time contours or templates of speaker specific acoustic features. An important element of these systems was the ability to time warp sample templates with model templates in order to provide useful comparisons. Most modern text dependent speaker verification systems are based on statistical representations of acoustic features analyzed as a function of time over specified utterances, most particularly the hidden markov model (HMM) representation. Modern text independent systems are based on vector quantization representations and, more recently, on Gaussian mixture model (GMM) representations. An important ingredient of statistically based systems is likelihood ratio decision techniques making use of speaker background models. Some recent research has shown how to extract higher level features based on speaking behavior and combine it with lower level, acoustic features for improved performance. The talk will present these topics in historical order showing the evolution of techniques.

**1pSC7. Fifty years of progress in speech and speaker recognition.** Sadaoki Furui (Dept. of Comput. Sci., Tokyo Inst. of Technol., 2-12-1 Ookayama, Meguro-ku, Tokyo, 152-8552 Japan)

Speech and speaker recognition technology has made very significant progress in the past 50 years. The progress can be summarized by the following changes: (1) from template matching to corpus-base statistical modeling, e.g., HMM and n-grams, (2) from filter bank/spectral resonance to Cepstral features (Cepstrum + DCepstrum + DDCEpstrum), (3) from heuristic time-normalization to DTW/DP matching, (4) from gdistanceh-based to likelihood-based methods, (5) from maximum likelihood to discriminative approach, e.g., MCE/GPD and MMI, (6) from isolated word to continuous speech recognition, (7) from small vocabulary to large vocabulary recognition, (8) from context-independent units to context-dependent units for recognition, (9) from clean speech to noisy/telephone speech recognition, (10) from single speaker to speaker-independent/adaptive recognition, (11) from monologue to dialogue/conversation recognition, (12) from read speech to spontaneous speech recognition, (13) from recognition to understanding, (14) from single-modality (audio signal only) to multi-modal (audio/visual) speech recognition, (15) from hardware recognizer to software recognizer, and (16) from no commercial application to many practical commercial applications. Most of these advances have taken place in both the fields of speech recognition and speaker recognition. The majority of technological changes have been directed toward the purpose of increasing robustness of recognition, including many other additional important techniques not noted above.

**1pSC8. Fifty years of progress in speech recognition.** Raj Reddy (Wean Hall 5325, School of Computer Sci., Carnegie Mellon Univ., Pittsburgh, PA 15213-3891, rr@cmu.edu)

Human level speech recognition has proved to be an elusive goal because of the many sources of variability that affect speech: from stationary and dynamic noise, microphone variability, and speaker variability to variability at phonetic, prosodic, and grammatical levels. Over the past 50 years, Jim Flanagan has been a continuous source of encouragement and inspiration to the speech recognition community. While early isolated word systems primarily used acoustic knowledge, systems in the 1970s found mechanisms to represent and utilize syntactic (e.g., information retrieval) and semantic knowledge (e.g., Chess) in speech recognition systems. As vocabularies became larger, leading to greater ambiguity and perplexity, we had to explore the use task specific and context specific knowledge to reduce the branching factors. As the need arose for systems that can be used by open populations using telephone quality speech, we developed learning techniques that use very large data sets and noise adaptation methods. We still have a long way to go before we can satisfactorily handle unrehearsed spontaneous speech, speech from non-native speakers, and dynamic learning of new words, phrases, and grammatical forms.

**1pSC9. Progress in speech research for telecommunications in the last five decades.** Biing Hwang Juang (School of Elec. and Comput. Eng., Georgia Inst. of Technol., Atlanta, GA 30332, juang@ece.gatech.edu)

Speech is the most natural form of human communications. A significant portion of speech research in the last century was devoted to the development of knowledge and technologies that aim at extending such a fundamental human capability over a long distance and at automating the support for such an extension. For the former, devices such as new acoustic transducers and techniques such as speech coding and echo cancellation have led to superb speech communication quality even for people located continents apart. For the latter, techniques in speech synthesis, recognition and understanding have been incorporated in various communication systems that automate billions of telephone calls or remote transactions of various sorts with remarkable success. Jim Flanagan, who had led the team on speech and acoustics in Bell Labs for over three decades, was very much in the center of this tremendous progress. In this talk, we highlight some of the milestones, many directly inspired by Jim Flanagan, that were achieved in the last few decades in the area of speech technologies and systems that have brought us benefits in our everyday life.

**1pSC10. Fifty years of progress in speech understanding systems.** Victor Zue (Comput. Sci. and Artificial Intelligence Lab., MIT, Cambridge, MA 02139)

Researchers working on human-machine interfaces realized nearly 50 years ago that automatic speech recognition (ASR) alone is not sufficient; one needs to impart linguistic knowledge to the system such that the signal could ultimately be understood. A speech understanding system combines speech recognition (i.e., the speech to symbols conversion) with natural language processing (i.e., the symbol to meaning transformation) to achieve understanding. Speech understanding research dates back to the DARPA Speech Understanding Project in the early 1970s. However, large-scale efforts only began in earnest in the late 1980s, with government research programs in the U.S. and Europe providing the impetus. This has resulted in many innovations including novel approaches to natural language understanding (NLU) for speech input, and integration techniques for ASR and NLU. In the past decade, speech understanding systems have become major building blocks of conversational interfaces that enable users to access and manage information using spoken dialogue, incorporating language generation, discourse modeling, dialogue management, and speech synthesis. Today, we are at the threshold of developing multimodal interfaces, augmenting sound with sight and touch. This talk will highlight past work and speculate on the future. [Work supported by an industrial consortium of the MIT Oxygen Project.]

**1pSC11. Microphones for speech and speech recognition.** James E. West (Electr. and Comput. Eng., Johns Hopkins Univ., 3400 N. Charles St., Baltimore, MD 21218)

Automatic speech recognition (ASR) requires about a 15- to 20-dB signal-to-noise ratio (S/N) for high accuracy even for small vocabulary systems. This S/N is generally achievable using a telephone handset in normal office or home environments. In the early 1990s ATT and the regional telephone companies began using speaker-independent ASR to replace several operator services. The variable distortion in the carbon microphone was not transparent and resulted in reduced ASR accuracy. The linear electret condenser microphone, common in most modern telephones, improved handset performance both in sound quality and ASR accuracy. Hands-free ASR in quiet conditions is a bit more complex because of the increased distance between the microphone and the speech source. Cardioid directional microphones offer some improvement in noisy locations when the noise and desired signals are spatially separated, but this is not the general case and the resulting S/N is not adequate for seamless machine translation. Higher-order directional microphones, when properly oriented with respect to the talker and noise, have shown good improvement over omnidirectional microphones. Some ASR results measured in simulated car noise will be presented.

**1pSC12. 50 years of progress in microphone arrays for speech processing.** Gary W. Elko (Avaya Labs, 233 Mt. Airy Rd., Basking Ridge, NJ 07920, gwe@ieee.org)

In the early 1980s, Jim Flanagan had a dream of covering the walls of a room with microphones. He occasionally referred to this concept as acoustic wallpaper. Being a new graduate in the field of acoustics and signal processing, it was fortunate that Bell Labs was looking for someone to investigate this area of microphone arrays for telecommunication. The job interview was exciting, with all of the big names in speech signal processing and acoustics sitting in the audience, many of whom were the authors of books and articles

that were seminal contributions to the fields of acoustics and signal processing. If there ever was an opportunity of a lifetime, this was it. Fortunately, some of the work had already begun, and Sessler and West had already laid the groundwork for directional electret microphones. This talk will describe some of the very early work done at Bell Labs on microphone arrays and reflect on some of the many systems, from large 400-element arrays, to small two-microphone arrays. These microphone array systems were built under Jim Flanagan's leadership in an attempt to realize his vision of seamless hands-free speech communication between people and the communication of people with machines.

NOTE: Separate registration required.

MONDAY EVENING, 15 NOVEMBER 2004

PACIFIC SALON 3, 7:00 TO 9:00 P.M.

**Session 1eID**

**Interdisciplinary: Tutorial Lecture on Ocean Noise and Marine Mammals**

George V. Frisk, Chair

*Department of Ocean Engineering, Florida Atlantic University, 101 North Beach Road,  
Dania Beach, Florida 33004-3023*

**Chair's Introduction—7:00**

***Invited Paper***

**7:05**

**1eID1. Ocean noise and marine mammals: A tutorial lecture.** Gerald D'Spain (Scripps Inst. of Oceanogr., UCSD, 9500 Gilman Dr., La Jolla, CA, 92093-0704, Mail Code: 0704, gld@mpl.ucsd.edu) and Douglas Wartzok (Florida Intl. Univ., Miami, FL 33199)

The effect of man-made sound on marine mammals has been surrounded by controversy over the past decade. Much of this controversy stems from our lack of knowledge of the effects of noise on marine life. Ocean sound is produced during activities of great benefit to humans: commerce, exploration for energy reserves, national defense, and the study of the ocean environment itself. However, some recent strandings of marine mammals have been associated with the occurrence of human-generated sound. The documented increase of man-made sound in the ocean suggests the potential for more extensive though subtler effects than those observed in the mass strandings. The purpose of this tutorial is to present the scientific issues pertaining to ocean noise and marine mammals. Basic physics of sound in the ocean and long term trends of ocean sound will be presented. The biology of marine mammals, particularly their production, reception and use of sound in monitoring their environment, social interactions, and echolocation, will be reviewed. This background information sets the stage for understanding the effects of man-made sound on marine mammals. The extensive gaps in current knowledge with respect to marine mammal distribution and behavioral and physiological responses to sound will highlight research needs.